EVIDENCE AND FORMAL MODELS IN THE LINGUISTIC SCIENCES

Carlos Gray Santana

A DISSERTATION

in

Philosophy

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

Michael Weisberg

Professor and Chair of Philosophy

Graduate Group Chairperson

Samuel Freeman, Avalon Professor of the Humanities and Graduate Chair of Philosophy

Dissertation Committee

Daniel J. Singer, Assistant Professor of Philosophy

Elisabeth Camp, Associate Professor of Philosophy, Rutgers University

EVIDENCE AND FORMAL MODELS IN THE LINGUISTIC SCIENCES

ABSTRACT

EVIDENCE AND FORMAL MODELS IN THE LINGUISTIC SCIENCES

Carlos Santana, Author

Michael Weisberg, Supervisor

This dissertation contains a collection of essays centered on the relationship between theoretical model-building and empirical evidence-gathering in linguistics and related language sciences. The first chapter sets the stage by demonstrating that the subject matter of linguistics is manifold, and contending that discussion of relationships between linguistic models, evidence, and language itself depends on the subject matter at hand. The second chapter defends a restrictive account of scientific evidence. I make use of this account in the third chapter, in which I argue that if my account of scientific evidence is correct, then linguistic intuitions do not generally qualify as scientific evidence. Drawing on both extant and original empirical work on linguistic intuitions, I explore the consequences of this conclusion for scientific practice. In the fourth and fifth chapters I examine two distinct ways in which theoretical models relate to the evidence. Chapter four looks at the way in which empirical evidence can support computer simulations in evolutionary linguistics by informing and constraining them. Chapter five, on the other hand, probes the limits of how models are constrained by the data, taking as a case study empirically-suspect but theoretically-useful intentionalist models of meaning.

TABLE OF CONTENTS

LIST OF FIGURES

PREFACE

Linguistics is hardly unique among the sciences in exhibiting a methodological (and sometimes sociological) divide between theoretical modeling and experimental evidence-gathering. The division of physicists into theorists and experimentalists is a familiar one, and the same distinction often pops up in other natural and social sciences. The case of theoretical and field ecology is a clear example, as is the case of neoclassical and behavioral economics. In some of these cases, such as the entrenched division of labor between mathematical theorizing and experimentation in physics, the two types of scientific activity have reasonably well-established means of mutually supporting each other in the quest to understand the world. This is not, unfortunately, always the case in linguistics.

Many linguists are fluent in multiple idioms of research and comfortably ecumenical in accepting the contributions of different approaches, but enough are not that regrettable divisions in field often crop up. These divisions are sometimes characterized by hot-blooded disputes, as is often the case between generativists and non-generativists. Sometimes these divisions are more of a quiet parting of ways. Work on parsing by computation linguists, for example, generally pays little attention to the work of formal syntacticians, and vice versa.

This dissertation doesn't offer a complete diagnosis of these and similar problems, nor does it prescribe a general-purpose treatment. It does, however, try to make progress towards understanding why and how experimental and formal approaches to studying language often fail to adequately inform each other. And in the three cases I look at—the preference for intuitions over other sorts of data in formal syntax and semantics, the hasty rejection of models of meaning that are contradicted by experiment, and the need for empirical grounding in simulating the evolution of language—I sketch some suggestions for how to treat some of the symptoms, at the very least. My hope is that this can help us not only improve our study of language, but also help us better understand how science works (and doesn't work) more generally.

CHAPTER 1: WHAT IS LANGUAGE?

Introduction

For the philosophy of linguistics, the question "What is language?" is the flipside of the question "What is the (proper) subject matter of linguistics?" Conceptual analysis for scientists has less to do with discovering concepts than with creating the concepts we need to meet our theoretical needs and reshaping them as needed to fit our growing understanding and theory. I am not primarily concerned with language, the folk concept, since while the folk concept may have been the jumping off point for research, it doesn't significantly constrain the science. So from the standpoint of philosophy of linguistics, to answer the question "What is the ontology of language?" we must begin by asking what sort of role the concept of language plays in linguistic theory and practice. It plays multiple roles, of course, but chief among these is that it picks out the object of study for linguists. Language, the scientific concept, is thus descriptively whatever it is that linguists take as their primary object of study, and normatively whatever it is they should be studying.

Regarding the descriptive question, I will argue that the object of linguistic study is multifaceted, comprising three separate but related types of entities. Many linguists take as their primary objects of study mental structures relating to language. The particular set of structures differs—a generative syntactician might take herself to only be studying language- (or even syntax-) specific structures, while many psycholinguists are happy to study any mental activity involved in linguistic processing—but everyone in this category takes their object of study to be psychological. Other linguists, especially those with ties to the social sciences such as sociolinguists, field linguists, and anthropological linguists, take their object of study to be primarily a social entity of some sort. Finally, some linguists take themselves to be studying abstract patterns evident in linguistic communication, with an ontology analogous to the metaphysics of mathematical entities. So in answer to the descriptive question "What is language?" we must respond that there are actually many types of language, roughly sortable in to three classes of ontologies, one psychological, one social, and one abstract.

To answer the normative question we look to see if there are reasons to favor one of these targets of inquiry over another. The most compelling reason to give up on one of the three facets of language would be to show that it either doesn't actually exist or that it is unsuitable for scientific study, and several philosophers of linguistics have tried to make just such a case. Partisans of one ontology over the others also appeal to answers to the descriptive question to answer the normative one, since extant scientific practice constrains to some extent which ontologies are legitimate. I'll review the most significant arguments of this sort, and give reason to reject them. Consequently, I'll conclude that we should give a plural answer to the question “what is the proper subject matter of linguistics?”

Although this analysis of the ontology of language preserves the extant diversity within the linguistic sciences, it does have some ramifications for the practice of linguistics. In particular, it tells against a tendency towards sub-disciplinary parochialism that is fueled in part by non-plural conceptions of the (proper) subject matter of linguistics. I'll also show how the fact of plurality requires making explicit the target of any particular work in linguistics. Agnosticism about the ontology of language should be avoided because he relation between hypothesis and evidence is shaped by the scientist's conception of her subject matter, and it follows that confirmation and theory choice can depend on which particular variety of “language” a researcher is studying.

Criteria for a conception of language

Debates over the proper subject matter of linguistics and correct ontology of language tend to come back to the same small set of issues, and by identifying these we can pick out the key criteria for a valid conception of language. Philosophers and linguists making a case that language is *x*, generally attempt to demonstrate three things:

(1) *x* exists, in a form accessible to scientific study,

(2) *x* is (descriptively) a primary object of study for linguists, and

(3) *x* is reasonably referred to as 'language'.

Additionally, most attempt to show that competing ontologies of language fail to satisfy one or more of these criteria. (1) through (3), then, appear to be taken more or less as necessary and sufficient criteria for establishing the proper subject matter of linguistics. Before moving on, then, let's take a closer look at each.

Advocates of none of the chief candidates for the ontology of language take language to be equivalent to the primary data gathered by linguists. Linguists, in the first place, study artifacts such as patterns of vibration in the air, symbols on a page, or reports of introspective judgments. Few linguists these days argue that language merely is these artifacts. Arguments that a preferred conception of language satisfies (1), then, typically appeal to an inference to the best explanation for observed patterns among these artifacts. *x* exists, the argument runs, because *x* is a theoretical posit licensed by the explanatory role it plays in our best theories explaining the primary data. Conversely, negative arguments against taking *y* to be the proper subject matter of linguistics can claim that positing *y* is unnecessary to explain the primary data, and thus *y* fails to meet (1). Of course, *y* can also fail to meet (1) if *y* isn't real, or if it is inaccessible to scientific inquiry.

As is the case with (1), partisans of all three camps sometimes argue that their ontology uniquely satisfies (2). For each of the types of ontology—psychological, social, and abstract—it isn't difficult to find linguists who hold it to be what they study. So non-pluralists must, and do, argue that to identify the primary object of study of linguistics we can't look to linguists' meta-theoretical reflections, but must instead infer it from their practice. Each camp, of course, claims that actual linguistic practice focuses overwhelmingly on their preferred ontology. I'll accept the premise that deeds, not words, determine whether *x* satisfies (2), since if I didn't, pluralism would be trivial to establish. Valid application of (2), however, takes more than an impressionistic sense about what day-to-day linguistic work allows us to infer about the primary object of linguistic research. I propose the following heuristic to determine what counts as a primary object of study in linguistic practice: if *x* is the common link between otherwise disparate objects of study, *x* is a good candidate for primary object of study. Suppose, for instance, that a linguist makes use of both data about subjects' eye movements and reaction times in a lab, as well as her own intuitions about semantic facts. Her intuitions and her subjects' eye movements have no direct connection, but both bear directly on language processing. Language processing is thus a good candidate for her primary object of study. This heuristic will allow us to assess arguments about whether a particular ontology satisfies (2).

I take (3) to be essentially practical. The issue is not so much that we need to hew closely to some prescriptively correct use of the term 'language,' but that only constrained disagreement about subject matter is possible within a research community. A particular scientist could come up with an idiosyncratic ontology which satisfies (1) and perhaps (2). For example, he might take language to be “information transferred through genetic material.” Now, information transferred through genetic material exists, and that scientist could certainly make it his primary object of study, but to call his subject matter 'language' in any way other than metaphorically would be problematic. Beyond the confusion that he would cause by doing so, it would insert him into the wrong research community. His work would have little to say to nearly all other linguists, and theirs would have little ramification for his. So to reasonably call an object 'language' in a scientific context, it must have at least some significant connection to what the community of linguists is already engaged in studying. This does not preclude novel uses of the word 'language' or novel conceptions of the ontology of language, but it does constrain which novelties are acceptable.

Having these criteria in hand allows us to situate the various arguments for one ontology of language over the others. Most such arguments will seek to establish that a particular account of a language satisfies all three, but its competitors do not. In what follows I'll defend the positive aspect of each argument—there are ontologies of all three classes which meet the criteria—but reject the negative by showing how the putative reasons to think the other ontologies fall short are misguided. We'll begin with the most-discussed type of linguistic ontology, treating language as an individual cognitive entity.

Language as psychological

The most influential advocate of language as a psychological entity is Chomsky, whose argument begins by establishing the same approach to metaphysics that we have adopted here. He argues that since the time of Descartes, it has been a common practice in philosophy to take the validity of the natural sciences as a fixed point, and metaphysics has reshaped itself around this fixed point (Chomsky 1995). The correct account of language, then, depends on the theoretical entities postulated by our best science, and according to Chomsky our best language science gives us an account of the state of the cognitive system responsible for language, which he calls both the 'language faculty' and “'I-language', 'I' to suggest 'internal', 'individual'” (1995: 13). Chomsky is aware that there is room for further specification. Does the language faculty include every physiological contributor to linguistic activity, including not just many parts of the brain but also parts of the vocal tract, etc.? Does the language faculty refer to the idiolect of speaker, meaning her unique, individual lexicon and grammar? Or should we take I-language to be more specifically the innate, universal biological endowment shared by speakers of different idiolects? Chomsky's own position is that language should be understood in the latter way; that is, as Universal Grammar, which consists of some minimal computational principles (Chomsky 2013). We need not follow him to such extremes, however, to accept the validity of a psychological account of the ontology of language. The cognitive sources of linguistic behavior exist and they seem to be a primary object of study for many linguists, so they satisfy two of our criteria for an ontology of language.

Dissension often focuses on the third criterion. Devitt and Sterelny (1989) call attention to the fact that Chomsky's idea of what language is departs significantly from what they call “Grandma's View.” Chomsky's definition of language would seem puzzling to Grandma, and Devitt and Sterelny think that Grandma is mostly right: linguistics “is about symbols and explains the properties in virtue of which symbols have their roles in our lives” (1989: 515). Since Chomsky is happy to write off the role language plays in communication as “peripheral” (2013: 655), he's not really using the term 'language' reasonably.

If this were the extent of the critique, it would have little bite. True, the technical definition of 'language' for Chomsky and his allies has little enough to do with the lay understanding of language or linguistics, but experts have been using the term in Chomsky's sense for half a century now, so we can't say it's unreasonable to call it 'language'. But Devitt and Sterelny's critique has more force than mere appeal to folk conceptions, since it is also meant to target our second criterion. Linguists themselves, they argue, are in practice much nearer to Grandma's View than Chomsky's, in that something like Grandma's View is the actual primary object of linguistics, despite what linguists may claim in their meta-theoretical reflections.

Devitt draws out this point clearly by an analogy (2003). Suppose we want to study horseshoes. We could gather samples of horseshoes and analyze them, or we could instead try to examine the psychological processes internal to the blacksmith when she creates horseshoes. Even if for some reason we did decide to approach the subject by looking at the blacksmith rather than the horseshoes themselves, we wouldn't think that her mental representations were the real horseshoes, and the shaped metal bars only objects of peripheral interest. Recall our rule of thumb for how to determine a primary object of study: if *x* is the common link between otherwise disparate objects of study, *x* is a good candidate for primary object of study. If you look at all the things involved in horseshoe-ology, you'd find not only the study of the blacksmith's expertise, but also study of the various uses horseshoes are put to (throwing implements in games, fashion accessories for horses, etc.), and of the symbolic roles they play (as lucky charms, as markers of cowboy culture, etc.). The common thread uniting all these objects of study is not the cognitive blueprint for horseshoes in the blacksmith's head, but the external, U-shaped bars of iron. So by our rule of thumb, the cognitive apparatus can't be the primary object of horseshoe-ology.

The horseshoes, obviously, correspond to Grandma's View of language, and the blacksmith's cognition to Chomsky's. Devitt's key point is not that to call the blacksmith's mental representations 'horseshoes' is intuitively silly, but that it doesn't correspond to what scientists would actually do. Linguists, Devitt argues, might pay lip service to the psychological account of language, but in practice they're actually concerned “with the properties of expressions in a language, symbols that are the outputs of a competence. [Their] work and talk seems to be concerned with items like the very words on this page” and not in the first instance with I-language (2003: 15). When we look at what linguists actually study, the common thread seems to be the external symbols used for communication. So Devitt argues that the psychological ontology of language not only fails the third criterion, but also falls short of the second, since it's not the primary object of study even for linguists who say that it is.

Devitt's arguments give us good reason to think that something like Grandma's View must be right as an ontology of language, but they don't suffice to disqualify the psychological ontology. He's right that much of linguistics is concerned with external symbols, but Chomsky's claim that the analysis of those external symbols—of performance, to use the technical term favored by generative linguists—can just be a method at understanding the underlying psychological phenomena is legitimate. After all, psychologists frequently use external measures to study internal psychological phenomena. If psychology consisted only of introspection and brain scans, we would have very little in the way of good psychological theory. When the psychologist asks a subject to adjust a patch of color until it matches another, he's studying visual cognition, not patches of color. When physicists observed tracks in cloud chambers, their primary object of study was subatomic particles, even though their data came in the first place from patterns of vapor. Along the same lines, Chomsky and his adherents seem to be justified in claiming that their primary object of study is a theoretical postulate which is observed only indirectly. Note the clear disanalogy with the horseshoe example: the varieties of data used by the psychologically-oriented linguist really are held together by the common thread of having a connection to the mechanisms of linguistic cognition.

To argue that the psychological ontology of language can't actually be the primary object of study, then, requires more than a claim that linguists often gather data that aren't psychological entities. We would have to show that those data do not serve as useful evidence for the psychological entities linguists often purport to study. Katz (1984) makes just such a case. He argues that the grammars produced by linguists accurately model performance, but there is little evidence that it captures the actual cognitive processes underlying linguistic behavior. This is in part because any particular pattern in performance could be produced by an infinite number of possible underlying cognitive systems, so we can't infer that the language faculty works in any particular way just from the primary data. It's difficult to come up with a straightforward rejoinder to Katz' argument, because it's no more than a special case of the thorny problem of underdetermination of theory by evidence. But this fact allows us an oblique response: while it's true that the primary data of linguistics are consistent with an infinite number of psychological grammars, this places psychologically-oriented linguistic theory in the same boat as all scientific theories, including those committed to different ontologies of language. So the issue Katz raises is one worth exploring[1], but the underdetermination of theory by evidence gives us no grounds to favor one subject matter of linguistics at the expense of the others.

Thus far we have seen how attempts to reject the claim that language is a mental entity on the basis of criteria (2) and (3) fail. Criterion (1), that the proper subject matter of linguistics must exist, gives even less ground for criticism. Claims that some particular account of the language faculty picks out a non-existent entity are legitimate, of course. It is an interesting and important question, for instance, whether Chomsky's "faculty of language, narrowly construed" (Hauser, Chomsky, and Fitch 2002) exists at all, and if it does exist, it is an interesting and important question whether it looks anything like what Chomsky says it does. But a negative answer to either of these questions is not a negative answer to the question of whether or not a psychology entity is a proper subject matter for linguistics. Even if no one yet has an accurate description of what the faculty of language is, there must be some cognitive facts about humans underlying our linguistic behavior and these facts describe the language faculty. So there must be some psychological ontology of language which satisfies (1).

In short, the language faculty must exist in some form or another, it is a primary object of study for many linguists, and given the history of modern linguistics it's reasonable for linguists to

[1] Presumably we can partially solve the problem by appeal to theoretical virtues such as simplicity, as well as by seeking the intersection of multiple forms of evidence.

call it 'language'. The psychological ontology of language is therefore a legitimate response to the questions “what is language?” and “what is the (proper) subject matter of linguistics?”

Language as social

It is not, however, the only legitimate response. Even Chomsky acknowledges, with rhetorical surprise, that other conceptions of language “remain current in contemporary cognitive science” (2013: 649). Many of those other conceptions treat language as a social object, and thus take linguistics to be at least partly a social science. As with the psychological ontology of language, the social ontology is a family of quite distinct conceptions of language rather than a uniquely defined object. While the ontology of social objects is a serious philosophical question[2], for present purposes we'll take it for granted that since the social sciences successfully study social objects, there is some correct account of social ontology. After all, we can give accurate descriptions and make successful predictions about social objects such as the Basque separatist movement, the song “Greensleeves,” this year's autumn fashions, and Oaxacan cuisine, so these must exist in some form or another. The question then becomes whether or not language can be properly understood as having an analogous social existence.

Many linguists and philosophers of language think it can, and they are generally lead to that conclusion by facts about language acquisition, methodological considerations, or by attending to facts about meaning. Labov (2012a: 6) argues that the object of study for sociolinguists must be a social entity because “we are programmed to learn to speak in ways that fit the general pattern of our community. What I, as a language learner, want to learn is not ‘my English’ or even ‘your English’ but the English language in general.” In other words, children learning language aren’t trying to learn any individual’s I-language, but trying to learn something that exists on a community level. Labov acknowledges that this is a contingent fact. We can conceive of language learners who did try to learn something which existed on an individual psychological level, such as a parent’s I-language. But all the evidence points to a social-level

[2] Some significant discussions: Durkheim (1895) Ch. 1, Searle (1995), Hacking (1999).

target for language learners, since “if they are brought into a new community before the age of nine, children will have the dialect system of that community, not of their parents” (2012a: 6). This fact gives us compelling reason to think that the scientific study of language acquisition needs a social ontology of language.

Linguists favoring a social ontology of language often do so because conceiving of language as a social object fits best with the types of questions they are interested in and the methodology and training standard in their sub-discipline. Clear examples are sociolinguistics and anthropological linguistics. Since these subfields take social sciences to be among their parent disciplines, both their research techniques and their guiding questions favor taking their object of study to have a social ontology. A popular sociolinguistics textbook, for instance, defines language as “what members of a society speak” (Wardhaugh 2006: 1) and emphasizes that “language is a communal possession” (2006: 2). Similarly, an upper-level textbook in linguistic anthropology situates the field as dealing with “language as a cultural resource and speaking as a cultural practice,” “language as a set of cultural practices,” and “language as a set of symbolic resources that enter the constitution of the social fabric” (Duranti 1997: 2-3). In fact, Labov, who has as much claim as anyone to have founded modern sociolinguistics, argues that “the central dogma of sociolinguistics” is that “the community is conceptually and analytically prior to the individual,” making language an “abstract pattern located in the speech community and exterior to the individual” (2012b: 266).

This is not to say that all sociolinguists or anthropological linguists take language to be a social object, only that language is so construed in the exemplars of those subfields. Nor do I mean to suggest that the social ontology of language belongs exclusive to subfields of linguistics with close ties to the social sciences. My point is simply that one good reason to adopt the social ontology instead of another option is because of the tools and interests of a particular research program.

Coming at the issue from a different direction, philosophers of various stripes have argued that language cannot be merely individual on the grounds that I-language is insufficient to account for facts about meaning. In particular, individual language users seem to attempt to conform to an external, social entity in their language use. The classic externalist arguments of Kripke (1972), Putnam (1975), and Burge (1979) can all be seen as arguing for the existence of social-type linguistic objects. Similarly, Dummett (1986) argues that we need a social ontology of language to explain how an individual speaker can be wrong about language. Even those of us who are skeptical of those philosophers' arguments for semantic externalism might be convinced by Lewis' (1969) treatment of language, which suggests that given its role in communication, language is best analyzed at the level of multi-agent interaction rather than at the level of purely individual psychology. So even if the social ontology of language supervenes on individual psychologies, these arguments give reason to think that the social ontology must play an important explanatory role in our language science.

So far then, language qua social entity seems to easily satisfy the three criteria we've set out. (1) It exists, in the same way that other typical objects of the social sciences exist. (2) It is a primary object of study for a significant number of linguists. (3) It is reasonably referred to as 'language'. In fact, it is probably the closest to both the everyday conception of language and the origins of linguistic study of the three types of ontology at question. For this reason, opponents of the social ontology attack it on the grounds than it fails (1) or (2), typically arguing that, construed as a social object, language does not exist in a form accessible to scientific study.

Unsurprisingly, the classic arguments that the social ontology fails (1) and (2) come from Chomsky. In Knowledge of Language (1986) he explains why it is a scientific mistake to take our object of study to be externalized language, or E-language. E-languages, he argues, “are not real world objects but are artificial, somewhat arbitrary, and perhaps not very interesting constructs” (1986: 26). He claims this in part on the grounds that folk understanding of language muddles sociopolitical facts with the linguistic facts, such as when we refer to 'Chinese' as a language despite Chinese dialects being as diverse as the Romance Languages. Additionally, the folk individuation of languages is incurably vague, as demonstrated by dialect continua such as the gradual geographic transition from German to Dutch. For these reasons, Chomsky argues, “all scientific approaches have simply abandoned [the sociopolitical] elements of what is called 'language' in common usage” (1986: 15). Consequently, a linguist can't just adopt the everyday concept of language, and must refine it into a technical notion of E-language. This refinement, however, necessarily involves idealization away from the facts of linguistic diversity. Even a refined technical concept of 'English' will need to elide many of the idiosyncratic differences between different speakers of English. So far, no problem. All science idealizes, so to reach the punchline of his argument Chomsky needs to further show why the idealizations leading to a technical notion of E-language are illicit.

To establish that the idealizations leading to E-language are problematic, Chomsky appeals to the unity of the sciences as a desideratum. “Linguistics, conceived as the study of I-language,” he proposes, “becomes part of psychology, ultimately biology” (1986: 27) This is desirable, he thinks, because the further we travel up the ladder of sciences, the closer we get to understanding the real mechanisms behind phenomena. “E-language, however construed, is further removed from mechanisms than I-language, (1986: 27)” so the idealizations behind the technical concept of E-language are taking us down the ladder of sciences—the wrong direction. Pulling Chomsky's argument together, we get the following: In idealizing from I-language to E-language, we lose access to the real mechanisms. We gain nothing from this loss, however, since the I-language conception can handle all the linguistic facts on its own. The social ontology of language thus fails to satisfy (1) because of both its “artificial nature” and its “apparent uselessness” (1986: 27). For similar reasons, it fails to satisfy (2). E-language is “an epiphenomenon at best” (1986: 25), so it can't be the primary object of linguistic study.

This is a sophisticated argument, but we should be skeptical of nearly all its premises. Categories which smoothly fade into each other, such as German and Dutch, can be real and useful categories. Wiggins (1997: 501) compares dialects to colors in this respect, and in general, so long as there are clear cases, the existence of borderline cases does not give compelling reason to be skeptical of a categorization scheme. If they did, all vague predicates would pick out entities inaccessible to scientific study. It isn't always the case that objects of scientific study need to be able to be outlined in terms of necessary and sufficient conditions, or even be clearly defined. The concept species remains ineliminable in biology despite all its conceptual problems, and a good case could be made[3] that languages, dialects, and so on are analogous to species in this respect. Linguistics can (and sometimes must) make use of fuzzy, ill-defined concepts just as other scientists do. The sociolinguistics textbook cited earlier points out, for instance, that “the concept [of speech community] has proved to be invaluable in sociolinguistic work in spite of a certain 'fuzziness' as to its precise characteristics” (Wardhaugh 2006: 119). It is significant that this is not an instrumentalist approach to sociolinguistic concepts; the author goes on to say that “speech communities, whatever they are, exist in a 'real' world” (2006: 120). Chomsky's criteria for proper scientific concepts seem to be too strict, and once relaxed they give no reason to exclude social conceptions of language.

Furthermore, Chomsky is also mistaken that all linguists have abandoned trying to define linguistic categories partially in terms of sociopolitical facts. Sociolinguists are often comfortable doing so, given their subject matter. Take, for example, Labov's seminal definition of 'speech community' which states that the concept is “not defined by any marked agreement in the use of language elements, so much as by participation in a set of shared norms” (1972: 120-1). A generative linguist, given her preoccupation with Universal Grammar, will want to avoid defining linguistic objects in terms of sociopolitical facts, but linguists with other scientific goals may have good reasons to use social facts to pin down their objects of study, including their primary objects of study. We should note as well that even in the counterfactual world where all linguists really did give up on defining languages (i.e. Urdu, Portuguese, Quiché) because they could not do so without appeal to sociopolitical facts, this doesn't mean we can't define language as a set of richly variegated social objects. So Chomsky's preliminary salvo on the social ontology of language—his attack on its resemblance to the folk concept—fails to hit the target.

[3] Lassiter (2008) makes just such a case.

His premises about scientific idealization are also fishy. Idealization is a property of how we represent our objects of study, not of the objects themselves, so idealized objects are not necessarily unreal. Chomsky himself must accept this, since I-language is constructed by idealization as much as E-language is. To construct the scientific object I-language, we must idealize away from individual variation, from cognitive limitations, and from the integration of the language faculty with other cognitive systems, to give a few examples. E-language concepts need not make the same idealizations but this gives us no reasons to favor E-language over I-language, nor vice versa. The idealizations made by linguists who adopt a social ontology and the idealizations made by linguists who adopt the psychological ontology are generally different, but neither is better in any strong sense. Given certain goals, certain idealizations are preferable, but linguistics need not be characterized by one true set of goals. We'll revisit this issue below, but for present purposes my point is merely that the idealizations involved in doing linguistics with E-language as primary subject matter do not necessarily make research more removed from the actual mechanisms than idealizations do in other approaches to linguistics.

Perhaps most perplexing of all of Chomsky's premises, however, is his claim that studying E-language is “useless” because we can handle all the linguistic facts by studying I-language. The only way such a claim is true is if we give 'linguistic fact' a narrow, ad hoc definition which excludes, say, facts about communication or language change. Chomsky and his allies do sometimes attempt just such a narrowing, and while they are justified in doing so for their own immediate research program, they have no grounds to suggest that their ad hoc narrowing is binding on the rest of the discipline. The fact of the matter is that many linguists pursue many questions which are best answered by appeal to facts about E-language, sometimes in tandem with facts about I-language. In summary, most of Chomsky's argument for why we shouldn't take social objects to be the target of linguistic inquiry falls apart under scrutiny.

Davidson (1986) attacks the existence of language from a somewhat different angle. He argues that “there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed” (1986: 265). The argument hinges on the potential for linguistic innovation. Language use involves error and creativity, and successful communication requires creating a hypothesis about meanings peculiar to each individual conversation. Any "language" stable enough to be studied would thus be too local to be of scientific interest. In short, we can't have scientific theories of English or Tagalog, because 'English' and 'Tagalog' pick out something different for each language-user at each conversational turn. We could have a scientific theory about the local language used by so-and-so and so-and-so in such-and-such particular interaction, but that wouldn't be an interesting or useful subject for science. For this reason, we might take studying language qua social object to be a doomed enterprise.

Davidson creates an excellent philosophical puzzle, but we need not infer from it the conclusion that social objects are an unfitting subject matter for linguistics. Instead of taking Davidson's philosophical puzzle as an impossibility theorem, we should take it as a scientific puzzle: given the fact of constant language change, how are communities of successfully communicating speakers maintained? This is a question answerable by language science. We can use sociolinguistic techniques to study the nature of language change and the constraints on linguistic novelty (e.g. Labov 2011), we can create formal models showing how speakers create a local language out of the resources of a broadly shared language (e.g. Cooper and Ranta 2008), or we could even, perhaps, do some philosophy of language (Armstrong 2016). Linguistic novelty, even frequent lexical innovation of the sort Davidson describes, is perfectly consistent with a shared language. Moreover, even if Davidson's puzzle gives some prima facie reason to be skeptical that languages exist, we can safely ignore that skepticism, just as we safely ignore philosophical skepticism in the sciences generally. Linguists and social scientists successfully create explanations, descriptions, and predictions in terms of shared public languages all the time. We can explain why I can successfully communicate with a typical Australian but not a typical Kazakh by appealing to the existence of English. We can describe the Great Vowel Shift as a historical change in the pronunciation of English. And we can do sensible sociological work about how second-generation immigrants to the United States almost always learn English.

Davidson's puzzle is not nearly compelling enough to force our science to abandon these sorts of claims.

This same point—that social objects like English and Tagalog play important roles in many of our best social scientific theories—also counters the other main argument against the social ontology of language. Some authors have suggested that even if languages qua social objects exist, they are superfluous from a scientific standpoint. Chomsky, for instance, argues that a “naturalistic approach to linguistic and mental aspects of the world seeks to construct intelligible explanatory theories, taking as 'real' what we are led to posit in this quest” (1995: 1). What we are led to posit, Chomsky thinks, is I-language and I-language alone, since we can explain all the linguistic facts by appeal to internal mental states. E-language “appears to play no role in the theory of language” (1986: 26) In a related vein, Heck argues that when it comes to the ontology of language, “the crucial question here is one of explanatory priority” (2006: 64), and clearly I-language is explanatorily prior, since all the social facts supervene on facts about individual speakers.

Neither of these arguments succeeds. Chomsky is right that in our quest for the best explanations, we are led to posit I-language, but we are also led to posit languages, topolects, dialects, speech communities, etc. This fact is undeniable, since these social objects are central posits in not only the work of many sociolinguists and field linguists, but also for generative syntacticians who discuss the features of individual languages, for computational linguists designing software meant to translate from one language to another, and for similar reasons in nearly every other subfield of linguistics as well. Chomsky's methodological naturalism certainly forces us to accept the psychological ontology as a proper subject matter for linguistics, but for the exact same reasons it forces use to accept the social ontology as well.

Heck's appeal to explanatory priority fails no better. He's likely correct about the metaphysical priority, since social facts are probably in-principle reducible (in some weak sense) to the psychological facts, but metaphysical priority doesn't entail explanatory priority. Explanations come in a variety of flavors (Salmon 1998; Lombrozo 2006), and linguistics can make use of explanations of psychological facts in terms of social facts as well as the other way around[4]. For instance, we appeal to differences in social linguistic environment to explain why language acquisition produces different results in different children. The question of explanatory priority will not favor one ontology of language over the other, and neither will reducibility. Even if we could in principle redescribe all the social facts as collections of psychological facts, in practice this would be a bad idea. Science might aim for unification, but it's clear that it doesn't aim for strong reduction. The goal of biology, for example, is not to eliminate talk of cells and species in favor of talk of atoms and molecules; biology instead embraces multiple levels of description, and there is no end goal to reduce them to one. Different scientific questions require treatment at different levels of granularity. Likewise, the goal of the social sciences, including many approaches to linguistics, is not to eliminate social objects in favor of talk of psychological facts. Linguistics has room to embrace multiple levels of description as well. In short, appeal to explanatory roles gives us more, not less, reason to accept the social ontology of language.

My arguments in this section have leaned heavily on the fact that linguists frequently appeal to social entities such as languages in their practice. The terms of the debate, accepted even by opponents of the social ontology such as Chomsky, are that scientific ontology is determined by the entities we are lead to postulate in the process of theory development. The standard objection to arguments such as mine is that reference to languages such as English or Southern Paiute is a matter of convenience, a useful set of fictions. Sometimes, I think, this is true, and a linguist referring to 'English' really just means a set of idiolects typified by a particular set of features. My claim stands, however, as longer as either (a) a significant number of linguists are not fictionalists about languages, or (b) the objects that 'English,' 'Southern Paiute,' and so on are convenient shorthand for are sometimes social objects. Both conditions are met if we consider the variety of linguistic disciplines, particularly those with close ties to the social sciences. In sum, given commonly accepted naturalistic methodology, (1) language has a real

[4] In fact, Lewis (1983) runs the explanatory priority argument in the other direction, against purely individualist conceptions of language.

social reality, and (2) language qua social object(s) is a primary object of study for a significant number of linguists. Since the social ontology never had a problem satisfying criterion (3), being a reasonable use of the term 'language', we can confidently conclude that the social ontology of language is a legitimate subject matter for linguistics.

Language as abstract

“Language as analogous to mathematics” might have been a more apt title for this section. Given that the ontology of abstract objects can be difficult to pin down, linguists who favor this third option for the ontology of language often argue for it on the grounds that we need an ontology for mathematical objects, and we can use that same ontology for language.

By way of background, we'll need to get some positions from the philosophy of mathematics on the table. *Platonism* takes a realist stance towards abstract objects. Numbers, functions, and so on exist abstractly in the same way that trees, stars, etc. exist concretely. Just as we perceive concrete objects, we intuit abstract objects. *Formalism* treats mathematics as a set of games according to which symbols are manipulated according to particular rules. Symbols need not be platonic abstract entities; they can be concreta such as scratches on a page, or they can be nominal categories. *Fictionalism* claims that mathematical entities have the same (un)reality as characters and objects in fictional stories. It is true that Hamlet murders Polonius, and false that Polonius murders Hamlet, but neither Hamlet nor Polonius is a real object. Likewise, it is true that 2 + 2 = 4, and false that 2 + 2 = 7, but '2' and '+' are not real objects.

Ontologies of all three sorts are in principle as available to philosophers of language as they are to philosophers of mathematics. Platonism is reasonably ascribed to some of the pioneers of formal semantics, such as Frege and Montague, but it has had its most thorough defender in Katz, who argues that “grammars are theories of the structure of sentences, conceived of as abstract objects in the way that Platonists in the philosophy of mathematics conceive of numbers” (1984: 18). I haven't found anyone who explicitly draws an analogy between the ontology of language and formalism in the philosophy of mathematics, but one plausible reading of the outdated school of American Structuralism could take formalism to be their position. Bloomfield's remarks that grammar is “the meaningful arrangement of forms in a language” (quoted in Chomsky 1986: 20) as well as the structuralist emphasis on the patterns of occurrence of discrete symbols both lend themselves to a formalist account of the subject matter of structural linguistics[5]. Finally, fictionalism in the linguistic context can be a thesis about the primary object of linguistic study. Consider, for instance, Carnap's methodological stance:

> “The direct analysis of [natural languages], which has been prevalent hitherto, must inevitably fail, just as a physicist would be frustrated were he from the outset to attempt to relate his laws to natural things—trees, stones, and so on. In the first place, the physicist relates his laws to the simplest of constructed forms; to a thin straight level, to a simple pendulum, to punctiform masses, etc.” (2002: 8).

To Carnap, the linguist, like the physicist, studies abstract models “in the first place.” Obviously, this methodological stance is consistent with the primary object of study for the scientist being the target system of his models, but it is also consistent with the models themselves becoming the primary object of study. The point of these examples is simply that the whatever our position for mathematical ontology, we could potentially adopt that same position for linguistic ontology and thus the abstract ontology can satisfy criterion (1).

The best argument for the abstract ontology is that some linguists are concerned merely with extensional adequacy. A grammar is extensionally adequate when correctly captures the linguistic facts of the natural language it is meant to describe. It must, for example, reject all the ungrammatical sentences of the language, and identify all the analytic sentences of the language (if any) as unconditionally true. Extensional adequacy does not require capturing the actual psychological procedure speakers use to arrive at the linguistic facts. In fact, as Katz (1984) observes, we can achieve extensional adequacy without any knowledge of the psychology facts, since there is always an infinite number of possible psychological procedures which could

[5] This strikes me as a more charitable understanding of the structuralist ontology of language than the crude behaviorism sometimes attributed to them.

produce the same set of linguistic facts. This doesn't mean that linguists can't look for the psychological facts, but it does mean that linguists can study the extensional linguistic facts without committing themselves to the psychological ontology. This occurs sometimes in sub-disciplines such as formal semantics, where some researchers will claim, for instance, that their analyses have “nothing whatsoever to do with what goes on in a person's head when he uses [the word in question]” (Dowty 1979: 375). This concern with merely the extensional facts, with disregard for the psychological and social aspects of language, marks linguists who are committed to an abstract ontology of some sort. Since such linguists exist, language, construed abstractly, appears to be a primary object of study for some linguists, so the abstract ontology meets criterion (2).

Chomsky's arguments for the non-existence of languages run as well against the abstract ontology as they do against the social ontology; in other words, not all that well. So instead of focusing on criterion (1) as a weak point of the abstract ontology, we'll consider criterion (2). The best argument against the abstract ontology is that it can't be a primary subject matter of linguistics because it's not worth studying. We have clear reasons for being interested in the psychological and social conceptions of language. Psychological and social objects have causal effects in the world, and linguistics construed as either psychology or a social science has clear connections to other sciences. Those virtues do not characterize linguistics construed as mathematics to the same extent. For that reason, Fodor argues, the primary problem with language qua abstract mathematical object is that “nobody is remotely interested in it” (1981). But we've seen that at least some linguists are interested in it, and I think they have good reason to be. Extensional adequacy is an easier target to hit than accurate psychological description, so for practical purposes where extensional adequacy is all that's required, it's beneficial to have linguistic theory aiming only at extensional adequacy. Consider, for instance, the applied linguistics involved in developing natural language processing systems. For most practical purposes, we don't care whether or not artificial intelligence systems use the same psychological mechanisms as humans do to generate linguistic behavior. The mark of a better natural language processor is greater extensional adequacy, not greater psychological accuracy. Since linguistics construed as mathematics is in some ways easier than linguistics construed as psychology, it's more likely to provide useful input certain applied settings, such as the development of natural language processing systems. In short, somebody *is* interested in language qua abstract object. We thus have reason to take the abstract ontology of language to be a potential primary object of study for linguists, even if it is often of less interest than the other two classes of ontology.

Why a plural ontology?

We've seen so far that all three classes of objects—psychological, social, and abstract—serve as satisfactory answers to the question, “what is the proper subject matter of linguistics?” In this section, I’ll argue that it follows that in the context of scientific linguistics, this implies a plural ontology of language. In other words, I’ll defend the methodological principle I introduced at the beginning: that in this context we can and should treat the questions of methodological subject matter and metaphysical status as equivalent.

This principle is restricted to the context of the scientific study of language. As I stipulated from the start, my talk of “ontology” refers to the ontology employed by the science, not necessarily to some fundamental accounting of what exists in what form in the universe. I’m agnostic about metaphysics in that broader sense. It may in fact be that in some extra-scientific metaphysical sense there is only one thing which is language, and the three classes of objects I’m discussing are its facets, parts, or different presentations. The possible existence of a unitary language in that sense, however, is compatible with the fact that the objects known as language in linguistics are plural. Linguists do not typically take as their object of study some chimeric entity combining the psychological, social, and abstract, either implicitly or explicitly. No one, so far as I can tell, has made a serious attempt to describe what such a metaphysically-odd chimera would be like, whereas there are many attempts to characterize what language in only one of the three categories would be like. To give a famous recent example, Hauser, Chomsky, and Fitch (2002) propose as scientifically useful two specifications of language qua psychological object. Linguists, in short, take as their ultimate objects of study language in one sense or another, not the language-chimera. This is not to say that, if the language-chimera actually exists, it may become an ultimate object of study in linguistics, but even if it did, unless it became *the only* ultimate object of study, pluralism would still be warranted.

My goal, it's worth keeping in mind, is to sort out disputes about whose approach to linguistics is correct, and these disputes take place in a discourse laden with ontological terminology: "this is what language *is,*" "languages don't *exist,*" etc. My pluralist solution takes this ontological terminology seriously—these really are discussions about the categories employed by and the objects studied by the science—and is thus able to engage with it on its own terms instead of summarily dismissing it as not even playing the right game. This is why it doesn't make sense for me to go in for a pluralism about methodology without a pluralism about ontology. To clarify, consider Sober and Wilson's discussion of a "pluralism of perspectives," where the same process (or in our case object) is fruitfully represented in different ways (1999: 331). Such pluralisms of perspectives do exist in linguistics, but the plural commitment to psychological, abstract, and social subject matters is not one of them. A linguist ultimately interested in how the brain parses speech and one interested in how languages change in contact with each other will have a lot to learn from each other, but they aren't merely taking two perspectives on the same subject matter. One is ultimately interested in a psychological subject matter, and her inquiry is satisfied once we understand the parser, even if we don't yet understand how languages change. The reverse is true for the other. If the difference between the two was merely one of perspective, they would, in the end, be after the same answers, but this is manifestly not the case. So insofar as we're discussing the status of language in scientific ontology, pluralism is the best way to make sense of practice in linguistics.

Two strictures on pluralism

My endorsement of pluralism about language isn't without restrictions. Pluralism has its pitfalls, so a sophisticated pluralism about the subject matter of linguists requires attention to two issues. On the one hand, sophisticated pluralism should not collapse into isolationism. On the other, it should forbid agnosticism.

Stricture 1: Pluralism should not collapse into isolationism.

The three classes of ontology we've considered bring along with them concomitant differences in subject matter and methodology, and are reflected to a large extent in the social structure of the discipline. It's tempting, given these differences, to postulate that what we really have are three Kuhnian paradigms. “There is a tendency,” the sociolinguist Labov observed back when *The Structure of Scientific Revolutions* was still fresh, “to see linguistics as a kind of debating society, where the winner is awarded the privilege of not reading the papers of the losers” because, after all, Kuhnian paradigms are incommensurable (1975: 56). So the social paradigm of language would have nothing to say to the psychological paradigm, and vice versa. We should not yield to this temptation, since, as Labov aptly observes, “the construction of such paradigms is a favorite occupation of those who would prefer to discuss the limits of knowledge rather than add to it” (1975: 56). The three ontologies of language are interrelated, and research bearing on one will usually have indirect significance for the others. If we treated linguistics as a number of distinct disciplines coincidentally located in the same academic department, we would miss out on knowledge to be gained through this indirect route.

Moreover, isolationism and insistence on incommensurability are antithetical to the important research project of understanding the connections between the three different types of object called 'language.' Questions about the metaphysical and causal relationships between the different ontologies of language are important questions for both linguists and philosophers of language. They are questions enabled by pluralism—we can't ask about the interaction of I-language and E-language, for instance, unless we accept both as real objects of study—but questions defeated by isolationism. So a sophisticated pluralism treats linguistics as a unified field characterized by multiple primary objects of study, not as three isolated fields with their own independent subject matters.

Stricture 2: Pluralism requires explicit commitment to an ontology

The second, equally great, danger of pluralism is the temptation for the individual scientist to be agnostic about the ontology of language. In any particular context, however, it is important to be explicitly committed to one or another primary object for a number of reasons. The first is that some debates are intractable unless it's clear what the subject matter is. Partee makes this point in her classic “Semantics—mathematics or psychology?” (1979). Semanticists, she points out, tend towards agnosticism about their subject matter and this leads to avoidable disputes. For example, she argues, Millianism about the semantics of names makes the most sense for certain extensional ontologies of language, but descriptivism works better for some psychological ontologies. The debate between Millianism and descriptivism is thus in part an artifact of the oft unacknowledged ontological commitments of the various discussants. This is a general phenomenon, because which claims about language are true depends on what language is. Consequently, when making a claim about language, it's important to be clear about which ontology is under discussion.

Similarly, commitment to an ontology needs to be explicit because the idealizations a scientist is licensed to commit depend on the subject matter at hand. Doing research into a particular linguistic question requires ignoring most of the facts about language, and distorting a few others to simplify the problem space. Whether it makes sense in a particular case to idealize away from a particular linguistic fact depends on which ontology we're trying to describe. For example, if we're currently taking the social ontology to be our primary subject matter, we might be comfortable ignoring individual performance errors, since they don't reflect the social facts of language. If we're after language construed as psychology, however, performance errors might be particularly illuminating data points, so we won't want to ignore them. Since the ontology of language bears on idealization, and idealization is central to scientific methodology, agnosticism about the ontology of language is untenable.

Finally, agnosticism about subject matter must be avoided because it makes confirmation difficult. No datum counts as evidence purely in virtue of its individual properties. Evidence is at least a two-place relation, in that all evidence must be evidence for something. If we're agnostic about a scientific subject matter, we're leaving out the second member of that evidence relation, so we can't be clear about the strength and valence of a particular piece of evidence. For example, the statistical analysis of linguistic data from Twitter speaks directly to language construed as a social object, but less directly to language processing. Conversely, data from fMRI brain scans might have more bearing on our understanding of the psychological ontology of language than on our understanding of language construed as a mathematical object. So when a linguist cites a data point as evidence in favor of a theoretical statement, we can't fully evaluate that claim unless we know which ontology of language she's committed to. This gives us yet another reason to avoid being agnostic or non-committal about the ontology of language.

References

Armstrong, J. (2016). The problem of lexical innovation. *Linguistics and Philosophy*, DOI: 10.1007/s10988-015-9185-9

Carnap, R. (2002). *The logical syntax of language.* Open Court Publishing.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use.* Greenwood Publishing Group.

Chomsky, N. (1995). Language and nature. *Mind*, 104(413): 1-61.

Chomsky, N. (2013). The Dewey Lectures 2013: What Kind of Creatures Are We? Lecture I: What Is Language?. *The Journal of Philosophy*, 110(12), 645-662.

Burge, T. (1979). Individualism and the Mental. *Midwest studies in philosophy*,4(1), 73-121.

Cooper, R., & Ranta, A. (2008). Natural languages as collections of resources. *Language in Flux: Relating Dialogue Coordination to Language Variation, Change and Evolution*. College Publications, London.

Davidson, D. (1986). A nice derangement of epitaphs. *Truth and Interpretation.* Blackwell

Devitt, M., & Sterelny, K. (1989). Linguistics: What's Wrong with" The Right View". *Philosophical perspectives*, 497-531.

Devitt, M. (2003). Linguistics is not psychology. In Alex Barber (ed.), *Epistemology of Language.* Oxford University Press.

Dowty, D. R. (1979). *Word meaning and Montague grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ (Vol. 7)*. Springer.

Dummett, M. (1986). A nice derangement of epitaphs: some comments on Davidson and Hacking. In Ernest LePore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson.* Cambridge: Blackwell

Duranti, A. (1997) *Linguistic Anthropology*. Cambridge University Press

Durkheim, E. (1895). *Rules of Sociological Method*.

Fodor, J. (1981). Some notes on what linguistics is about. *Readings in the Philosophy of Psychology*, 2, 197-207.

Hacking, I. (1999). *The social construction of what?*. Harvard University Press.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve?. *Science*, 298(5598), 1569-1579.

Heck, R. (2006). Idiolects. In Judith Jarvis Thomson & Alex Byrne (eds.), *Content and Modality: Themes from the Philosophy of Robert Stalnaker*. Oxford University Press

Katz, J. J. (1984). An outline of Platonist grammar. *Talking Minds: the study of language in the cognitive sciences*, 17-48. MIT Press

Kripke, S. A. (1972). *Naming and necessity.* Springer

Labov, W. (1972). *Sociolinguistic Patterns*. University of Pennsylvania

Press

Labov, W. (1975). *What is a linguistic fact?* Humanities Press International.

Labov, W. (2011*). Principles of Linguistic Change, Cognitive and Cultural Factors (Vol. 3)*. John Wiley & Sons.

Labov, W. (2012a). *Dialect diversity in America: The politics of language change*. University of Virginia Press.

Labov, W. (2012b). What is to be learned? The community as the focus of social cognition. *Review of Cognitive Linguistics* 10(2): 265-293.

Lassiter, D. (2008). Semantic externalism, language variation, and sociolinguistic accommodation. *Mind & Language*, 23(5), 607-633.

Lewis, D. (1969). *Convention: a philosophical study*. John Wiley and Sons.

Lewis, D. (1983). Languages and Language, in *Philosophical Papers Vol. 1*. Oxford University Press

Lombrozo, T. (2006). The structure and function of explanations. *Trends in cognitive sciences*, 10(10), 464-470.

Partee, B. (1979). Semantics—mathematics or psychology?. In *Semantics from different points of view* (pp. 1-14). Springer

Putnam, H. (1975). The meaning of ‘meaning’. *Minnesota Studies in the Philosophy of Science* 7:131-193

Salmon, W. C. (1998). *Causality and explanation.* Oxford University Press.

Searle, J. R. (1995). *The construction of social reality*. Simon and Schuster.

Sober, E., & Wilson, D. S. (1999). *Unto others: The evolution and psychology of unselfish behavior*. Harvard University Press.

Wardhaugh, R. (2006) *An introduction to Sociolinguistics, 5th edition.* Blackwell

Wiggins, D. (1997). Languages as social objects. *The Journal of Philosophy*, 72(282), 499-524.

CHAPTER 2: NOT ALL EVIDENCE IS SCIENTIFIC EVIDENCE

A puzzle about scientific epistemology

Norms of modern scientific practice include a standards of evidence much stricter than the standards of evidence for everyday inference. My aim is to justify these stricter norms, even though they exclude some otherwise legitimate evidence from scientific consideration. Before getting into *why* not all evidence is scientific evidence, let's consider a few vignettes to convince ourselves that not all evidence is scientific evidence.

Vignette 1a:

Quinn gets a nasty case of food poisoning. She calls her sister, who recommends boiled cloves. 'I was skeptical when I first heard about it,' her sister says, 'but I tried it and it did wonders for nausea.' Even though she's not fully convinced it will be effective, on the basis of her sister's experience Quinn gives the folk remedy a try.

Vignette 1b:

Quinn is a pharmacologist trying to discover an agent that will reduce symptoms of nausea. She calls her sister, who recommends boiled cloves. Quinn writes up a proposal for a NIH grant, noting that "We have some preliminary evidence that cloves reduce nausea, in the form of my sister's experience." After getting funding, and performing randomized controlled trials, Quinn submits an application for FDA approval to market cloves as an anti-nausea agent. On the application she cites all her evidence, including both the outcome of the RCTs, and her sister's anecdote.

It would be surprising to see an appeal to one individual's experience on a biomedical document. Nevertheless, it seems perfectly reasonable for Quinn to allow her sister's testimony to increase her confidence that cloves are an efficacious remedy. The example works as well with conventional wisdom substituted for anecdote: Quinn knows that it is conventional wisdom that cloves fight nausea, etc. In individual decision making, we rely on common sense and conventional wisdom all the time, and this seems appropriate. Even though conventional wisdom is often wrong, following its dictates is probably more reliable than choosing randomly. Nevertheless, we would be surprised to see explicit appeals to conventional wisdom in most scientific writing.

Vignette 2a: Tevita is planning a vacation to the Amazonian Andes. He wants to get along with the locals, so he chats with a friend who had visited Peru the year before, and asks questions about the Quechua people. On the basis of what his friend says, Tevita forms several beliefs about Quechua customs.

Vignette 2b: Tevita is an anthropologist, writing an ethnography of the Quechua people. Money's tight, so rather than doing field research himself, Tevita interviews American tourists returning from Peru about Quechua customs. He uses these interviews as the primary data for his ethnography.

Tevita the scientist's actions seem wrong, while Tevita the tourist seems justified. Testimony is an ineliminable source of evidence in everyday life, but it seems that scientific norms exclude the testimony of non-experts in certain circumstances[6].

Vignette 3a:

Mariana is working on her calculus homework. She has little tolerance for working through problem sets but she's mathematically gifted. She often looks at a problem and just has a hunch

[6] Note that who counts as an expert is determined by the question we're asking, not merely by possession of some credential. The testimony of American tourists is not scientific evidence for an ethnography of the Quechua, but it is evidence for an ethnography of American tourism.

about the answer, and these hunches are almost always right. She uses these hunches to fill out her homework instead of working through the problems.

Vignette 3b:

Mariana is a gifted mathematician. She often has hunches of the form "*p* is a theorem of formal system *q*." It's well known that her hunches are generally accurate. Since this is the case, and since she hates coming up with formal proofs, her published papers are all of the form "I have a hunch that *p*, therefore *p*."

In personal decision making it makes a lot of sense to rely on hunches, intuitions, etc. These data are rarely acceptable as evidence in scientific discourse, however. They may play other roles, such as helping the scientist generate hypotheses, but do not generally count as evidence.

We could adduce examples from nearly every domain of science. Unlike the lay public, linguists don't consult high school English textbooks to determine grammaticality, geographers don't use automobile odometers to determine distance, and meteorologists don't rely on old sailor's rhymes to predict tomorrow's weather. Nevertheless, each of these sources of evidence seems safe for personal use.

These vignettes give us a feel for the puzzle we're going to tackle. In each of these cases, a kind of data which constitutes satisfactory, though perhaps weak, evidence for an individual epistemic agent does not constitute satisfactory evidence for the joint epistemic project of science[7]. So, in asking "why isn't all evidence scientific evidence?" we're looking for the features of scientific epistemology which make it relevantly distinct from individual epistemology.

[7] Note that this issue here is stronger than the problem, raised by Achenstein (1983, 2001), that not everything which raises the probability of a hypothesis intuitively counts as evidence. In the cases I'm raising, it does make sense to call the data evidence, just not in scientific contexts.

Since I'm claiming that not all evidence in the strict sense counts as evidence in science, we need a definition of scientific evidence.

SCIENTIFIC EVIDENCE: *E* is scientific evidence for a hypothesis *H* iff *E* is evidence[8] and *E* is acceptable as confirmatory in scientific practice.

I intend ‘confirmatory’ to mean incremental, not just absolute confirmation (cf. Carnap 1962), and it includes evidence which disconfirms as well as that which confirms *H*. The word 'acceptable' in this definition should be taken as normative, meaning that *E* is scientific evidence only if it's the kind of thing scientists should accept as evidence. With this definition in hand, we can see that evidence in the strict sense and scientific evidence are conceptually distinct, since a datum can be technically confirmatory without being acceptable as confirmatory in scientific practice. The vignettes illustrate how they are frequently actually distinct. Contemporary scientific practice excludes some data which is evidence in the strict sense. Anecdata, common sense, hunches, non-expert testimony, outdated measurement techniques, high school textbooks and so on are often evidence in the strict sense, but not scientific evidence.

One obvious reason why not all evidence is scientific evidence is that scientists have limited resources. If we only have the time or funding to pursue a limited number of observations, it makes sense to pursue the most epistemically valuable data. Weaker evidence might end up not being scientific evidence because we don't have the resources to gather it under scientific conditions. Most exclusions of evidence in the strict sense from scientific argumentation, however, can't be chalked up to limited resources, because in most cases the weaker types of evidence are already available or obtainable at very low cost. The scientist often already knows

[8] I’m deliberately non-committal on what it is to be evidence in the strict sense. I assume that any reasonable account of evidence (that evidence is the thing which justifies believe, that evidence = knowledge, that evidence is that which should alter credences, etc.) will suffice to generate the puzzle, because whichever of these definitions we accept, some evidence in the strict sense will be excluded by contemporary scientific practice.

the conventional wisdom or anecdata that apply to the question at hand, and intuition-pumping, consulting non-expert opinion, and so on are all nearly costless means of data gathering. If anything, the effect of limited resources would be to favor weak evidence, so we can't explain why not all evidence is scientific evidence by a mere appeal to limited resources.

To make the puzzle even more precise, here’s one more definition:

PRINCIPLE OF TOTAL EVIDENCE (PTE): In decision making and hypothesis confirmation take into account all available relevant evidence.

The name comes from Carnap (1947), but the idea is an old one. Just to be clear, PTE doesn't say that before you make a judgment you must go out into the world and track down every piece of relevant data first. It merely recommends using all the data you have available at the time of judgment, as well as any that can be obtained costlessly. Carnap justifies PTE in part by noting that it is "generally recognized" and that it would be "obviously wrong" to violate it (1947, 138-139), and certainly PTE is quite intuitive. I. J. Good, however, proposes that PTE is not supported by intuition alone, arguing that PTE follows from "the principle of rationality—the recommendation to maximize expected utility" (1967, 319). Good makes his case using a mathematical proof, but the basic idea is understandable qualitatively. When making a decision, taking into account an additional piece of evidence can only increase the objective likelihood of achieving the best available outcome. So as long as using or obtaining an additional piece of evidence is costless, it is always beneficial from the standpoint of utility to do so. Since an individual's total evidence includes only that which they already possess, and evidence already-in-hand is costless, to maximize expected utility an individual must follow PTE.

If you don't think that questions of epistemic rationality can be treated by appeal to practical rationality and expected utility, you'll need grounds other than Good's to accept the principle of total evidence. Perhaps Carnap's appeal to its intuitiveness works for you; perhaps you have your own reasons. Whatever your reasons, if you accept PTE[9], we've come to the puzzle. PTE states that epistemic rationality demands that we take into account all our evidence. But in scientific practice we don’t do so (or at least pretend not to), and in fact we seem to think it is epistemically wrong to do so. So either scientific practice egregiously violates a straightforward epistemic principle, or PTE doesn't apply in scientific contexts. I favor the latter for two reasons. First, PTE doesn't apply to scientists because it is a principle for ideally epistemically rational agents, and scientists are not much more ideally rational than the rest of us. Second, PTE is meant to apply to individual agents, but scientific inquiry is inherently social.

These two facts mean that the epistemic ends of science are better served by a different principle of evidence, which we'll call the scientific standard of evidence. Different kinds of evidence vary in what I’ll term *reliability*, where a token piece of evidence from a more reliable type of evidence is less likely to be misleading than a token piece from a less reliable type. The scientific standard of evidence states that, roughly, only the most reliable sources of evidence are acceptable as scientific evidence. We can make this more explicit.

SCIENTIFIC STANDARD OF EVIDENCE (SSE): A type of evidence is acceptable scientific evidence if either

(a) It is highly reliable, or

(b) It is among the most reliable types of evidence available.

Let's call condition (a) the absolute criterion and (b) the relative criterion.

The idea behind the absolute criterion is that we don't always want to exclude very good evidence just because nearly impeccable evidence is available. What counts as highly reliable probably varies a bit from discipline to discipline, but all that matters is that a vague threshold for

[9] If you reject PTE, perhaps you accept a near neighbor which yields the same puzzle. If you reject anything in the ballpark, then there is no puzzle, and you can treat the puzzle-solving I'm about to engage in as further arguments for a claim you already accept—that PTE is false.

absolute reliability exists. The relative criterion exists because for some scientific questions good evidence is hard to come by. This doesn't mean we give up on asking those questions, it just means that we have to make do with the best evidence available. Having a relative criterion as a disjunct to the absolute criterion ensures that we can do science even when the only evidence available is poor.

A few further comments on SSE. First, SSE is an aspirational norm and mechanisms for enforcing it are often informal, imprecise, and indirect, so we should expect to see imperfect adherence to it. Occasional violations of SSE don’t demonstrate its inexistence. It does seem, however, that SSE accurately describes the standard of evidence across a wide array of sciences—consider the variety among the vignettes with which we began. Second, successful application of SSE presupposes that we have at least a rough idea of how reliable a type of evidence is[10]. So where the level of reliability is in question, SSE predicts that we should expect scientists to fight over whether a certain type of evidence is acceptable. Likewise, SSE explains in part why reproducibility is so important—it helps establish the level of reliability of a type of evidence.

Having specified SSE, we can restate the puzzle as "Why does science use SSE rather than PTE?" The answer, as I've suggested, is that science is a collective endeavor of non-ideal agents. To demonstrate, I'll compare what happens if science follows SSE to what happens if it follows PTE, showing that in many common situations the epistemically superior outcome comes from following the stricter criteria in SSE. I'll use case studies and simulations to show that these counter-intuitive results fall out of scientists' bounded rationality and the social nature of science. Showing that SSE leads to superior epistemic outcomes than PTE is sufficient to both justify and explain the scientific departures from PTE. It shows, in short, that treating relatively weak evidence as counting as evidence in science will frustrate the goals of science. We thus have

[10] Several commentators have suggested to me an even stronger prerequisite—that we know why the type of evidence in question is reliable. I think this is descriptively inaccurate, but won't argue the point here, since nothing in this paper hangs on whether or not we accept this stronger requirement.

pragmatic reason to think of evidence in a different way in the scientific domain than in other domains.

Evidence and Incentive

Let's idealize for the moment and stipulate that the goal of science is to come up with true theories. The first reason to prefer SSE to PTE is that the goal of the individual scientist is different. Although most scientists aim to produce true theories, they also have other ends, including receiving credit for their work and using their time efficiently.

The chief method by which a scientist gets credit is by publishing research. Generally, these publications take the form of a set of claims and the evidential support for this set of claims. Consider a scientist aiming to publish a paper she has written arguing for hypothesis *H*. Imagine that the norms of her discipline accept two sorts of evidence: *e*, which is weak but easy to obtain, and *E* which is strong but costly to obtain. Knowing that her paper will be accepted even if containing only evidence of type *e*, she will likely only try to obtain evidence of type *e*. Her peers will generally do the same, because it’s the most efficient means to gain credit. In the aggregate, this leads to suboptimal epistemic outcomes, since the better sort of evidence *E* will be underutilized in the discipline.

Note that the problem is less severe in the case of the individual agent. An individual’s decision about whether or not to pursue *e* or *E* will be determined by weighing the cost of obtaining *E* against the benefit it provides for their expected utility. If obtaining *E* is worthwhile, the rational agent will do so. In the case of a group of scientists, however, even if all the scientists are ideally rational they may all individually choose *e* because it dominates choosing *E* from the individual perspective. This of course undermines the goal of their scientific discipline. Depending on how much the scientists themselves value obtaining true theories, it may also be sub-optimal for their own utility, in a sort of social dilemma.

Let me give two examples. I'm hesitant to make critical claims about a specific scientific discipline without detailed, careful argumentation, so both examples will be of scientists criticizing their own discipline on the grounds that their colleagues have flocked to weak evidence *e* at the cost of strong evidence *E*.

Our first example comes from generative syntax. Wasow, a syntactician, and Arnold, a psycholinguist, make the following claims about evidence in generative syntax: "standards of data collection and analysis that are taken for granted in neighboring fields are widely ignored by many linguists. In particular, intuitions have been tacitly granted a privileged position in generative grammar. The result has been the construction of elaborate theoretical edifices supported by disturbingly shaky empirical evidence" (2005: 1481-82). This is a polemical claim, of course, so Wasow and Arnold back it up with a detailed argument. Their argument, in fact, proceeds just as we would expect it to if SSE were taken as normative in linguistic research. First they outline the varieties of evidence available to the syntactician, including intuitions, corpus data, and psychological experiments. They then give evidence that intuitions are markedly less reliable as evidence than the other two sources. Intuitions, to use our present terminology, are *e*, while experimental and corpus data are *E*. Among the evidence they cite for this claim are empirical results showing that intuition data exhibits much more variability than usage data. Moreover, they illustrate with a number of examples that when a linguist's intuition runs counter to usage data, we reject the intuition in favor of the usage data, which shows that usage data is accepted as the more reliable sort of evidence. Their examples are also meant to show that usage data contradict intuitions fairly frequently, so intuitions must not be highly reliable in an absolute sense. Taken together, these observations show that intuitions are (a) not highly reliable, and (b) significantly less reliable than feasible alternative sources of evidence. Wasow and Arnold note that despite this fact, most syntax papers appeal largely or only to intuitions as evidence, and they argue that this must be harmful to the discipline as a whole.

Let's assume that Wasow and Arnold are correct in their assessment of the uses of evidence in generative syntax. The problem seems to be that intuitions are both easy to obtain and considered satisfactory evidence by the field, so linguists are disincentivized from pursuing stronger evidence. This, unfortunately, diminishes the epistemic quality of the output of the field.

Were the field to hew more closely to SSE, however, intuition data would no longer be sufficient to support a theoretical claim, so linguists would have to rely on stronger forms of evidence. What I'm suggesting, then, is that if Wasow and Arnold are right, generative syntax could improve the epistemic quality of its output by adhering to SSE. This seems to be what Wasow and Arnold think as well: they conclude by arguing that "linguistic inquiry should be subject to the methodological constraints typical of all scientific work" (2005, 1495). SSE, I'm arguing, is a reasonable account of those pan-scientific methodological constraints. If syntacticians accepted SSE, however, it would require rejecting PTE, since intuition-data is virtually costless[11]. In short, the case of generative syntax shows that the fact that scientists seek easy credit requires us to reject PTE in favor of SSE.

As similar example in a different discipline comes from the provocatively titled "Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior?" (Baumeister, Vohs, and Funder 2007). Consider two types of evidence available to social psychologists, direct observation of the behavior in question and self-reports. Call the first behavioral methods, and the second survey methods. Behavioral methods give us significantly more reliable evidence, so behavioral methods give us evidence *E*, while survey methods give us evidence *e*. This is not to say that surveys and self-reports are no better than uninformed guessing. They often are, and so constitute evidence in the strict sense, but they are much less reliable than behavioral methods. Unfortunately, they are also much cheaper. Surveys take less time, both to design and deploy, and are often easier to analyze as well. So if my point about incentive for easy credit leading to poor epistemic outcomes is true, we would expect a disproportionate number of social psychologists to be publishing survey evidence instead of

[11] Note that if intuition data were not virtually costless, then PTE would be compatible with Wasow and Arnold's conclusion, because PTE does not tell against prioritizing which methods we use to gather evidence. Even, however, in cases where weaker evidence is not virtually costless, a flat-out prohibition on weak evidence (SSE) will lead to the epistemically preferable outcome because it precludes the possibility of rationalizing the pursuit of weak evidence at the expense of strong evidence.

behavioral evidence. According to Baumeister, Vohs, and Funder we do indeed find this to be the case. They recount that

> "personality psychology has long relied heavily on questionnaires in lieu of behavioral observation, a state of affairs that has begun to change only recently and ever so slowly, at that. Even worse, social psychology has actually moved in the opposite direction. At one time focused on direct observations of behaviors that were both fascinating and important—a focus that attracted many researchers to the field in the first place—social psychology has turned in recent years to the study of reaction times and questionnaire responses. These techniques, which promised to help to explain behavior, appear instead to have largely supplanted it. The result is that current research in social and personality psychology pays remarkably little attention to the important things that people do" (2007, 396).

Of course, the authors give evidence for this claim. They examined the two most recent issues of the reputable Journal of Personality and Social Psychology. In the most recent issue, they identified 38 published studies. Of the 38, only one involved direct observation of behavior, and that one they classified as borderline. The previous issue, however, was 100% better. Out of 38 studies in that issue, two involved direct behavioral evidence.

Note that the problem isn't that questionnaires aren't evidence. The authors of this complaint acknowledge that "Self-reports are often illuminating" especially because in some cases, such as studies of emotional experience, they "are all that is possible" (2007: 399). In other words, when it comes to questions where the relative criterion obtains—where self-reports are the best option despite being unreliable in an absolute sense—psychologists should use self-reports. This stance coheres with SSE. No, the problem is that on all the questions where the relative criterion doesn't obtain the best available evidence is being neglected because lesser evidence is good enough to get credit. The solution, again, is to enforce something like SSE. In an ideal world where scientists were incentivized merely to provide the best epistemic contribution possible, they would find the best balance of gathering both behavioral and survey data. But scientists are not so incentivized. If we were to forbid appeal to self-reports when better data is available, it would lead to the use of better evidence and thus truer theories in social psychology. Self-reports would join anecdata and appeals to conventional wisdom in the bin of weak evidence not worth our time. By adopting SSE, we would improve the epistemic quality of science.

I don't mean to beat up on social psychology or generative syntax. Both fields produce interesting results, and many researchers do avail themselves of the best techniques. The point is merely that leading researchers in both fields have identified a problem caused by credit-seeking and claim it could be resolved by adherence to something like SSE. I don't mean to beat up on credit-seeking either. It plays an ineliminable role in science, since we want motivated scientists. SSE, unlike PTE, allows us to both encourage credit-seeking and mitigate the potential it has to epistemically undermine scientific practice. So one reason for favoring SSE, attested in at least two contemporary disciplines, is that it helps ensure that the best sources of evidence are not neglected.

Cognitive bias and evidence-gathering

A second set of reasons why scientists should avoid PTE comes from the fact that scientists are as subject to cognitive biases as the rest of us. PTE assumes ideal epistemic agents, but these biases mean that none of us is ideal in the appropriate sense[12]. Consequently, following PTE can actually lead to worse epistemic outcomes in a number of circumstances. In this section, I'll review some well-demonstrated traits of human reasoning, and show how given these traits, following SSE would lead to epistemically superior outcomes.

[12] Ideal, that is, in the sense of being the sort of agent in idealized decision-theoretic models. It has been argued on empirical grounds (Gigerenzer and Brighton 2009) that given our actual environment, such agents are not actually optimal reasoners. If this is the case, then the 'biases' discussed in this section are in fact good reasoning strategies for individual learners and the problems I attribute to them will emerge particularly in the context of joint scientific research.

The first problem caused by cognitive biases has to do with the timeframe of evidence acquisition. Remember that at issue is whether scientists should ignore evidence that they get for free—in other words, the type of evidence at issue is the kind the scientist will already have been exposed to at the beginning of inquiry. If they had not already been exposed to the evidence, they would have to seek it out, so it wouldn't be free, and PTE wouldn't apply. Many of the examples of questionable scientific evidence are of this costless, already-in-hand sort. Anecdotal evidence, intuitions, conventional wisdom, and so on all fall into this category. Furthermore, the types of data that don't are usually easier to gather, so if a scientist is going to gather multiple kinds of data, they usually start with the weaker but easier to obtain evidence, such as survey data.

The problem with this is that a slew of well-demonstrated cognitive biases show that there is a first-mover advantage in evidence assessment. One set of these biases has to do with overrating the first bit of evidence you receive. The bias that Kahneman and Tversky (1974) identify as the anchoring effect, for instance, is our tendency to do precisely that. They give as a simple example of anchoring the following experiment: two groups of high school students were given five seconds to estimate an arithmetic expression. One group was given 8 × 7 × 6 × 5 × 4 × 3 × 2 × 1, and the other given 1 × 2 × 3 × 4 × 5 × 6 × 7 × 8. The two expressions both multiply to 40,320, but the median estimate for the first group was 2250, contrasting with a median of 512 for the students given the ascending sequence. Which evidence the students processed first, that is, had a disproportionate impact on their assessment.

The anchoring effect is similar to another cognitive bias called the focusing illusion. Focusing occurs when any piece of information salient in the immediate context is taken to be especially relevant to a judgment, whether it actually is or not. Strack et al. (1988) provide a clear experimental example of focusing. They asked subjects two questions:

(1) "How happy are you with life in general?" and

(2) "How many dates did you have last month?"

Asked in that order, answers to the two questions showed no significant correlation. When (2) preceded (1), however, correlation rose to 0.66, a very significant correlation for psychology. In other words, merely making a certain set of otherwise irrelevant facts salient made those facts dominate subjects' judgments. Both anchoring and focusing have been demonstrated repeatedly and shown to apply in a number of different situations (Kahneman et al. 2006). Furthermore, subjects do not appear to be able to avoid the biases even if they are made aware of them (Wilson et al. 1996) and have difficulty resisting the pull of anchors even when given monetary incentive to avoid them (Simmons et al. 2010).

So between anchoring and focusing we have two well-demonstrated and recalcitrant cognitive biases which make it so that we will tend to be disproportionately influenced by the first piece of evidence we attend to. Why is this a problem? For the following reason. These biases mean that we're going to overrate evidence no matter what. The question is how to mitigate the effect of this overrating. If we're going to take a piece of evidence to be more decisive than it actually is, better to go with a highly reliable datum than a fairly unreliable one. Therefore, overrating weak evidence is more harmful than overrating strong evidence.

Suppose we use PTE in science. Then the first-mover advantage goes to the weak evidence. Suppose we really internalize SSE, however. The scientist will dismiss the weak evidence, and the first salient evidence will be the strong evidence. So with PTE we overrate and focus on weak evidence, but with SSE we overrate and focus on strong evidence. This gives us one reason to prefer SSE to PTE.

The first-mover problem caused by cognitive bias occurs in a different way as well. Suppose for the sake of argument that we have a scientist who successfully suppresses anchoring and related biases, and thus accurately updates her belief when presented with weak evidence. Will taking weak evidence into account still impede her scientific research? Yes, in many cases it will, because of motivated reasoning in the form of confirmation bias and research bias.

Confirmation bias, perhaps the most discussed, experimented-upon, and well-demonstrated human cognitive bias, is "unwitting selectivity in the acquisition and use of evidence" to support a favored hypothesis (Nickerson 1998, 175). This occurs through favoring confirmatory evidence, ignoring alternative explanations, looking only for positive cases, and over-weighting confirmatory evidence (Nickerson 1998). Related is research bias, also called experimenter bias, which is a catch-all term referring to the subtle and unintentional ways in which researchers' prior beliefs lead them to manipulate the outcome of experiments. Philosophers are familiar with research bias in the famous example of Clever Hans, a horse supposedly able to perform arithmetic, but who was actually responding to unconscious bodily cues from its trainer. Clever Hans appeared to be able to do arithmetic only because its trainer believed that it could and thus unwittingly manipulated a demonstration of that fact. A more apposite and contemporary example is the frequency of p-hacking in psychological research. P-hacking occurs when psychologists take advantage of "researcher degrees of freedom" to make what should be a null result come out as a positive one (Simmons, Nelson, and Simonsohn 2011). For example, a researcher might add subjects until they find a positive result, then end the experiment. Similarly, they might (unconsciously) choose which statistical methods to use, variables to analyze, or comparisons to make in order to increase their chances of confirming their hypothesis. Research bias in the guise of p-hacking appears to be ubiquitous in published psychological research (Simonsohn, Nelson, and Simmons 2014), and since there's nothing special about the brains of psychologists, we should expect research bias to be common in other disciplines as well.

My intent in bringing up confirmation bias and research bias is not to cast doubt on scientific results, but to highlight a pitfall in research—a pitfall that can be exacerbated by adherence to PTE. Scientists following PTE will begin by consulting weak evidence, and update credence in their hypotheses on the basis of this weak evidence. Since the weak evidence likely agrees with their prior hypotheses, their credence in those hypotheses will go up. Likewise, if a scientist is open-minded, consulting weak evidence will shift him from his neutral stance. In either case, the consequence of consulting weak evidence is that the scientist is more likely than before to engage in confirmation-bias-induced shenanigans, tainting the epistemic outcome of their research. A scientist who adheres to SSE, on the other hand, will not have this increased chance of unwittingly manipulating results.

SSE then, can lead to better outcomes than PTE because it mitigates some of the problems caused by confirmation and research bias. We should note, however, that this is not the case if the epistemic disvalue of the bias is outweighed by the epistemic contribution of the weak evidence. But this is generally going to be consistent with SSE. Recall that SSE doesn't proscribe comparatively weaker evidence if that evidence is fairly strong in its own right. Cases where the epistemic value of the weaker evidence trumps the negative effects of bias will probably be cases where the comparatively weaker evidence is strong enough to pass SSE. In other words, in either possible case, SSE performs as well as or better than PTE.

One response to this argument would be to argue that cognitive biases in science are too trivial to worry about. Because the errors introduced through biases such as confirmation bias are generally small and subtle, we might think that they are generally drowned out by good scientific methodology. If this were the case, we could safely ignore the cognitive biases of scientists. Unfortunately, this is not the case. Mathematical modeling shows that even small amounts of bias have a major deleterious effect on the percentage of scientific claims which are true (Ioannidis 2005). Additionally, most professional incentives in science probably increase confirmation bias (Nosek et al. 2012), meaning that the bias is unlikely to be small in the first place. Strict standards of evidence are thus necessary to mitigate some of this bias.

To recap, scientists are human, and they must contend with the effects of cognitive bias on their research. PTE does nothing to correct for these biases, but the hard line drawn by SSE does diminish their effect in common cases. This gives us additional reason to favor SSE, and perhaps an additional explanation for why scientists treat weak evidence as if it were not evidence at all.

Evidence and the social structure of science

A third reason for favoring SSE over PTE comes from the collective nature of the scientific enterprise. In brief, the problem is that in situations where an individual determines their beliefs in part by taking the beliefs of their peers as evidence, weak sources of evidence can lead to a harmful false consensus. This is because if a number of researchers respond to weak evidence, it might create a quick consensus in the discipline, and this consensus can be taken to be strong evidence that a hypothesis is true. For example, suppose astronomers attend to the fact that folk wisdom suggests that the moon is made of cheese. Further suppose that each individual astronomer recognizes that folk wisdom is only weak evidence, and thus comes to believe the hypothesis that the moon is made of cheese only tentatively. So far so good; everything each astronomer has done so far is reasonable. At this point, however, each will notice that the hypothesis is nearly universally held to be true among their peers, but it will not always be plain that a peer's belief is weak or that it is based on weak evidence. In virtue of this, there will appear to be a consensus in the discipline that the moon actually is made of cheese, and this will solidify each astronomer's belief in the hypothesis. But if the folk wisdom serving as evidence has the potential to be systematically misleading, then this solidified consensus is unwarranted.

This is an instance of what economists call an information cascade (Banerjee 1992). In situations where you have reason to believe that others may have information you lack, it can be rational to follow the crowd, even if your private information contradicts the crowd's behavior. Behaviors can therefore spread through a population principally on the basis of their popularity, as has been confirmed in a number of empirical domains. Voters, for instance are known to be influenced by opinion polls, and marketers know that they can create real demand for a product by buying up product to make it appear that such demand already exists (Bikhchandani et al. 1998). Although using popular behavior as a source of information in this way is often a sensible move from an individual perspective, it can easily lead to suboptimal group outcomes (Banerjee 1992), as in the moon-cheese case.

The reason the possibility of information cascades favors SSE over PTE is that information cascades are only likely to lead to false consensus if the evidence underlying individuals' original behavior is relatively weak. False consensuses are a particularly bad outcome for a scientific discipline. Not only can a false consensus directly obscure the truth, but future work premised on a false theory will also be misleading to the community. Additionally, false consensuses in science often prove extraordinarily difficult to dislodge. If only strong evidence is allowed in a reasonably large scientific community, however, information cascades will almost never result in false consensus. To demonstrate this, I developed an agent-based computer simulation. The simulation shows how allowing weak evidence in science leads to a risk of an information cascade to a false conclusion—a risk not present with strong evidence, or in the case of an individual reasoner. This favors SSE over PTE as the scientific evidentiary standard. We turn now to the details of the simulation.

Overview

The purpose of the model is to explore how the effect of evidence on scientists' beliefs is affected by the social structure of science. I designed the model in NetLogo, software designed for developing agent-based models. The model has only one kind of agent, representing researchers. Researchers consist of a CREDENCE[13] in the hypothesis in question, a BELIEF calculated from that credence, and a set of one-directional outgoing LINKS to other researchers, which represents the set of other scientists whose opinions they respect. Additionally, a global variable EVIDENCE-STRENGTH represents the strength of the kind of evidence available to all researchers, and another global variable HYPOTHESIS-IS-TRUE represents whether the hypothesis in question is true or not.

Initialization and processes

[13] Disclaimer: the names of simulation parameters should not be taken too seriously. BELIEF, for instance, should not be equated with belief, where belief is our best philosophical account of belief. Likewise for CREDENCE, EVIDENCE-STRENGTH, etc. These parameters are merely highly idealized representations used to model epistemic agents.

When the simulation is initialized, the set of researchers is created. Each consecutive researcher is assigned a set of LINKS to agents chosen randomly from the set of agents, weighted by the number of ingoing LINKS each agent already possesses plus one. For example, an agent with three in-LINKS is four times as likely to be selected as one with no in- LINKS. This yields a scale-free network[14], which is the type of network exhibited by citation patterns in the sciences (de Solla Price 1965). Finally, researchers are randomly assigned an initial CREDENCE between 0.3 and 0.7.

The model has two steps. First, each researcher individually assesses evidence. In each individual case, the valence of the evidence is determined probabilistically based on EVIDENCE-STRENGTH. For example, if EVIDENCE-STRENGTH is 0.6, each researcher has a 60% chance of finding evidence which agrees with HYPOTHESIS-IS-TRUE, and a 40% chance of finding misleading evidence. Each researcher then updates their CREDENCE on this evidence by Bayesian conditionalization. After updating, each researcher recalculates their BELIEF, believing the hypothesis only if their CREDENCE is greater than 0.5.

Second, each researcher observes the BELIEF of researchers it has outgoing LINKS to, and adjusts its CREDENCE on the basis of what it perceives the general attitude to be. Specifically, the researcher determines the percentage of peers who believe the hypothesis, then takes the midpoint between that percentage and its present CREDENCE to determine its new CREDENCE. For example, if an agent has a CREDENCE of 0.4 and 2 out of 10 peers believe the hypothesis, the agent will adopt a new CREDENCE of 0.3.

I have two comments on this second process, since this is where the effects of the social network come into play. First, that researchers only attend to their peers' beliefs and not their credences is justified because while we have at least rough first-person access to our own degrees of belief, we generally don't have precise third-person access to the degrees of belief of

[14] The same simulation run on random networks produces similar results.

others. We have excellent third-person access, though, to the polar beliefs of others, and the model design reflects these facts.

Second, the formula I use to model how researchers take into account the beliefs of peers is admittedly simplistic. Starting simple, however, has been a fruitful approach for researchers modeling the social structure of science. Drawing on Kitcher’s (1993) seminal work on formally modelling the social structure of science, a number of philosophers of science have used agent-based models to understand joint epistemic activity. The simplifications in my model resemble those in models prominent in the literature. There is precedent, for instance, in having agents assess how many of their peers have polar belief in a hypothesis (Zollman 2010). And Grim et al. (2011) model social influence on belief merely by having the agent average the credences of all its neighbors and adopt that average as its own credence. While even more simple than the analogous mechanism in my model, their formula captures enough of the phenomenon to make the result plausible, but is simple enough that it's clear why the result occurs. In making extraordinary simplifications, I am following the lead of what has been a fruitful methodology, as well as yielding to necessity. The psychological mechanisms of resolving near-peer disagreement are not yet well understood, so keeping the model simple allows us to aim for rough similarity with a broad range of plausible psychological mechanisms. In particular, the method I use has several features which are both realistic-looking and supported by experimental work on attitude change (Petty and Wegener 1998): agents take into account particularly the opinions of those they regard as reliable, agents weight their own opinion more heavily than that of any other individual, and agents care more about the general consensus than the testimony of individuals. So although the formula I use is simplistic, the results may still be illuminating.

Run parameters

For the purpose of analysis, I ran the simulation 60,000 times. This included runs with EVIDENCE-STRENGTH set at 0.55, 0.6, 0.65, 0.7, 0.75, and 0.8 for 1000 runs at each value with HYPOTHESIS-IS-TRUE set to true, and an additional 1000 runs at each value with HYPOTHESIS-IS-TRUE set to false, for a total of 2000 runs at each value. This was repeated five times, with each researcher having outgoing LINKS to 5, 10, 15, 20, and 25 others in each respective repetition. Each run included 200 researchers[15].

Output and analysis

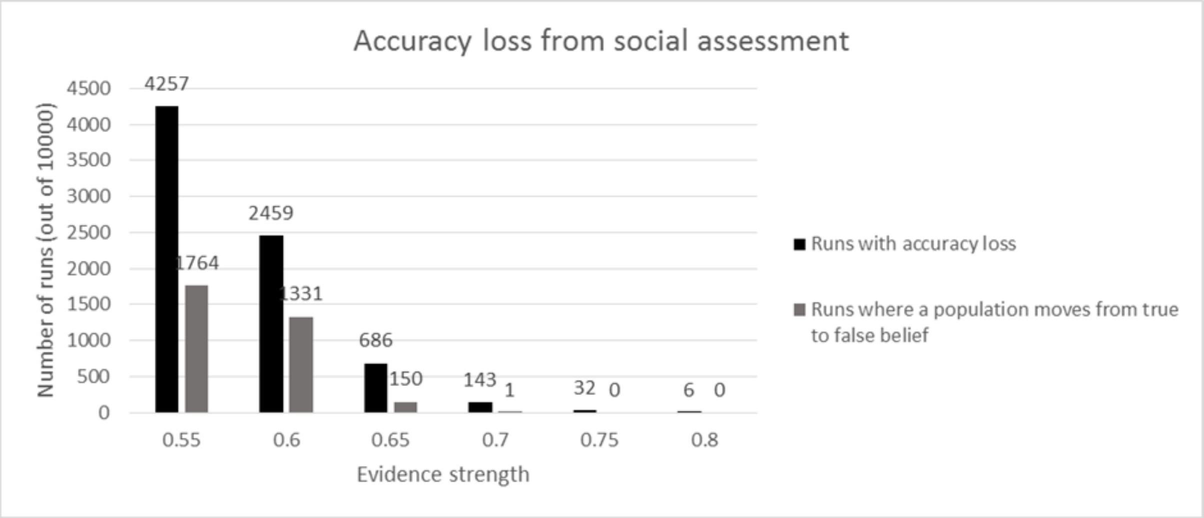
Since the purpose of this simulation is to determine the interaction between evidence strength and the social nature of scientific reasoning, the primary output measure, PROPORTION-CORRECT, is the proportion of the research population with a true belief. For instance, if the hypothesis is false and 60 out of 200 agents believe it is false, then PROPORTION-CORRECT would be 0.3. For each run of the simulation, I had NetLogo report PROPORTION-CORRECT after setup (pregame score), after researchers update on the evidence (halftime score), and after researchers adjust their beliefs based on their social network (final score). This allows us to differentiate between the effects of initial evidence assessment and the resolution of peer disagreement, which is necessary because we are particularly interested in the latter.

The effect of initial evidence assessment is precisely what we would expect: There is some chance that evidence will mislead individuals when it is weak, and we see that here. 2711 out of 10,000 runs at evidence strength 0.55 saw a drop between pregame score and halftime score, indicating an increase in false belief across the population, as did 649 at strength 0.60 and 38 at strength 0.65. There was never an increase in false belief with stronger levels of evidence. This doesn't tell against PTE, since despite this risk of error with weak evidence the expected

[15] More than one reader has asked about whether the results reported below are robust to increasing the level at which a CREDENCE is treated as a BELIEF. In response, I ran the same set of simulations with BELIEF requiring a CREDENCE of at least 0.55, and again with the same parameter set at 0.60. The outcome differed from what is reported below in two respects. First, the number or runs where a population moves from true to false belief was lower at low levels of evidence strength, and tapered off more slowly as evidence strength increased (there were still no runs where this occurred at the highest levels of evidence strength). This difference was more pronounced at 0.60 than at 0.55. Second, the number of runs with accuracy loss due to social assessment increases significantly at 0.55, and even more so at 0.60. At low levels of evidence strength, it occurs the majority of time when the hypothesis in question is true. This appears to be because the more stringent criterion for belief leads to an appearance of less general confidence in the hypothesis. The effect, however, diminishes rapidly at higher levels of evidence strength. This is a different social epistemological benefit of excluding weak evidence than the one I draw on below, but either way excluding weak evidence has a notable quasi-qualitative effect.

epistemic payoff is positive. I report these results only for purposes of comparison with the data of interest.

The real question is what happens when researchers pay attention to their social environment. Does resolving disagreement at low evidence strength lead to epistemically detrimental information cascades? In other words, do a significant number of runs increase in false belief between halftime and the final score? Yes:



At evidence strength 0.55, 4257—well over a third of runs— show an increase in false belief when researchers take the beliefs of their peers into account. 2459 runs have the same result at evidence strength 0.60, and at higher levels of evidence strength the number of runs with a loss in true belief falls off sharply.

The key scenario, however, is when the social network itself creates a false consensus—when a population whose members mostly hold the true belief is converted into a population with a majority of members holding the false belief. At the lower levels of evidence strength, 0.55 and 0.6, this occurred in 1764 and 1331 out of 10,000 runs, respectively. At the highest levels of evidence strength, social assessment was never able to overturn a general acceptance of the truth. The opposite effect—a population with generally false belief converting to a population with generally true belief—is less frequent, occurring in 1296 runs at evidence strength 0.55, and 382 runs at evidence strength 0.6.

What does this mean? It doesn't mean, on its own, that we should ban evidence of strength 0.60[16] or lower from scientific practice, since despite the high rate of error it still might maximize expected epistemic payoff to use the weak evidence. But it does mean—and this is all I intend the model to demonstrate—that weak evidence does frequently lead to quick, but inaccurate consensus. The simulation suggests that scenarios such as the moon-cheese fable will be more than rare oddities. Such scenarios are bad news for science not because they don't maximize expected epistemic accuracy in the short term, but because they put a halt to progress on questions that should remain open. One of the lessons of modeling the social structure of science has been that in the long run, quick consensus can be hard to displace and tends to be less accurate than the results of long-term debate (Zollman 2007; Grim et al. 2011). I'm making a similar, but distinct point. The short term maximization of epistemic accuracy obtained by utilizing low quality evidence isn't worth it, I argue, because those of cases where a quick false consensus emerges are too heavy a cost to pay. A false consensus generally impedes scientific progress more than a leaving a question open.

The negative consequences of researchers accepting a hypothesis as accepted by consensus are severe. Not only does the field largely stop looking for what is in fact the right answer to that particular research question, but even worse, that false claim can become an accepted presupposition for future theorizing and thus skew the interpretation of future results. This latter effect particularly gives us good reason to want to prevent any significant number of false consensuses in science. What we learn from the simulation is that false consensuses as a result of the social structure of science are a problem mostly when our primary sources of

[16] I don't think we should attack any special significance to these numbers in particular. It would be a mistake, for example, to reformulate the absolute criterion in SSE to specifically draw the line at 60% reliability. The mistake would lie not only in taking the details of this particular model too seriously, but also in forgetting that SSE is the result of multiple factors, not just the one the model is designed to identify.

evidence are weak, so this gives us a reason to favor a standard of evidence which forbids the use of weak evidence. In other words, it gives us yet one more reason to prefer SSE to PTE.

Further discussion

We began with a puzzling fact. Not everything which would be confirmatory to an ideally rational being is accepted as confirmatory in science. I presented this puzzle as a competition between two norms: the principle of total evidence (PTE), which states that rational confirmation uses all available evidence, and the scientific standard of evidence (SSE), which states that scientific confirmation uses sources of evidence which are among the best available.

I've attempted to give a partial answer as to why the scientific community is rationally justified in adhering to SSE rather than PTE. I first argued that by forbidding the use of weaker sorts of evidence we better incentivize individual scientists to make optimal contributions to joint epistemic projects. In support of this reasoning, I presented two case studies of contemporary sciences where leading figures have identified major problems in their fields resulting from too lax a standard of evidence. I then argued that PTE is only optimal for ideal agents, and gave examples from experimental psychology of systematic biases in human reasoning which entail that SSE yields better epistemic returns for actual human cognitive agents. Finally, I argued that because confirmation in science is inherently social, weaker types of evidence can lead to harmful false consensus.

I make no claim that these explanations are exhaustive, but I do think that they account for a significant portion of why something like SSE is accepted in most scientific disciplines. Some readers may worry that it isn't true that something like SSE is accepted across the sciences. It doesn't take a professional sociologist to notice that as a matter of fact, however much scientists might profess adherence to something like SSE, scientific theorizing and argumentation frequently does involve invocation of weaker sorts of evidence such as intuition and anecdote. I have two responses to this objection. First, even deeply rooted norms are not observed perfectly. Even though modern science generally does accept something like SSE, we should still expect to find violations. But we should also find that those violations, if noticed, are policed from within the discipline. This is precisely what's going in in the case studies from generative linguistics and social psychology. Second, SSE is only about what is permissible for use in confirmation. Weaker sorts of evidence may still play important roles in the context of scientific discovery, such as in helping scientists formulate hypotheses, design experiments, and create tentative theories. Given that discovery is as much a part of science as is justification, we should expect to see things like anecdotes and intuition-data appearing in scientific practice.

I'll close by suggesting some implications of this account of why not all evidence is scientific evidence. First, while SSE is not a solution to the problem of demarcation between science and pseudoscience, it does provide a heuristic that can help with some of the questions an account of demarcation is meant to answer. For example, in scientific settings it justifies us in ignoring claims of alternative medicines if they are based entirely on weaker sorts of evidence. In fact, SSE indicates that it would be wrong to take seriously those claims or the evidence supporting them.

A second implication of my arguments is that other epistemic contexts may require a stricter standard of evidence. The obvious example is the courtroom, where in many nations certain types of evidence are already excluded. My treatment of the scientific standard of evidence in terms of purely epistemic reasons for adopting it could be adapted to the legal context, though non-epistemic values may also need to be accounted for. We could also give similar treatments to other institutional and organizational epistemic contexts, such as policy-making, intelligence-gathering, and market research. Furthermore, I think some of the reasons I've outlined here apply even to individual contexts, and while I won't argue for it here, there's an argument to be made for why not all evidence is evidence for individual human beings. I'm content for the time being, however, to have shown merely how scientific reasoning benefits from its higher standard of evidence.

References

Achinstein, P. (1995) "Are Empirical Evidence Claims A Priori?" *British Journal for the Philosophy of Science* 46: 447-73.

Achinstein, P. (2001) *The Book of Evidence*. Oxford University Press.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 797-817.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *The Journal of Economic Perspectives*, 151-170.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior?. *Perspectives on Psychological Science*, 2(4), 396-403.

Carnap, R. (1947). On the application of inductive logic. *Philosophy and phenomenological research*, 8(1), 133-148.

Carnap, R. (1962), *Logical Foundations of Probability*. 2nd ed. Chicago: University of Chicago Press.

de Solla Price, D. J. (1965). Networks of Scientific Papers. *Science*, 149(3683), 510-515.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1(1), 107-143.

Good, I. J. (1967). On the principle of total evidence*. British Journal for the Philosophy of Science*, 319-321.

Grim, P., Singer, D. J., Reade, C., & Fisher, S. (2011). Information Dynamics Across Sub-Networks: Germs, Genes, and Memes. In *AAAI Fall Symposium: Complex Adaptive Systems*.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2006). Would you be happier if you were richer? A focusing illusion. *Science*, 312 (5782), 1908-1910.

Kitcher, P. (1993). *The Advancement of Science: Science Without Legend, Objectivity Without Illusions.* Oxford University Press.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7(6), 615-631

Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 323-390). New York: McGraw-Hill.

Pfungst, O. (1911). *Clever Hans:(the horse of Mr. Von Osten.) a contribution to experimental animal and human psychology*. Holt, Rinehart and Winston.

Simmons, J. P., LeBoeuf, R. A., & Nelson, L. D. (2010). The effect of accuracy motivation on anchoring and adjustment: do people adjust from provided anchors?. *Journal of personality and social psychology*, 99(6), 917.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359-1366.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer*. Journal of Experimental Psychology: General*, 143(2), 534.

Strack, F., Martin, L. L., & Schwarz, N. (1988). Priming and communication: Social determinants of information use in judgments of life satisfaction. *European journal of social psychology*, 18 (5), 429-442.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185 (4157), 1124-1131.

Wasow, T., & Arnold, J. (2005). Intuitions in linguistic argumentation. *Lingua*, 115(11), 1481-1496.

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, 125(4), 387.

Zollman, K. J. (2007). The communication structure of epistemic communities. *Philosophy of Science*, 74(5), 574-587.

Zollman, K. J. S. (2010). Social structure and the effects of conformity. *Synthese*, 172(3), 317-340.

CHAPTER 3: LINGUISTIC INTUITIONS ARE NOT SCIENTIFIC EVIDENCE

"[M]any of the philosophical ripostes to generative linguistics misfire because they fail to incorporate Chomsky's fundamental methodological precept that linguistics is a science and not an a priori discipline"

—John Collins (2008, 24)

Linguistics is a science and not an a priori discipline. This presents us with a trilemma:

(1) If a discipline relies on intuition as a primary source of evidence, it is an a priori discipline, and not a science.

(2) Some subfields of linguistics—notably generative syntax and formal semantics[17]—rely on intuition as a primary source of evidence.

(3) Linguistics is a science and not an a priori discipline.

We could escape the trilemma by rejecting (3), but this would be to commit the philosopher's mistake that Collins admonishes us to avoid. A response friendlier to linguistics is to reject (1) on the grounds that some intuitions count as empirical, scientific evidence. Linguistic intuitions, this argument goes, are distinctive among intuitions in possessing this empirical quality. Thus at least one discipline, linguistics, can rely on intuitions as evidence and yet not be an a priori discipline.

In this chapter I present the leading accounts of how linguistic intuitions might constitute empirical, scientific evidence, and provide reason to reject them. Given the failure of these

[17] Hereafter just "syntax" and "semantics." I make this concession to brevity apologetically, acknowledging the diversity of approaches to syntax and semantics falling outside the present discussion.

accounts, I conclude that linguistic intuitions are not scientific evidence. This fact requires an alternative escape from the trilemma, and I argue for rejecting (2). Contrary to appearances, I contend, intuitions do not necessarily function as evidence in syntax and semantics, thus those fields can be safely counted among the empirical sciences.

Sources of evidence in syntax and semantics

The evidential status of a type of scientific data depends on the varieties of relevant data available. Our discussion of the evidential status of linguistic intuitions should thus be prefaced with an overview of alternative sources of evidence.

In addition to intuitive judgments, linguists have at their disposal a wide array of sources of evidence for syntactic and semantic features of language. A typical catalog of sources of evidence in linguistics comes from Krifka (2011), who usefully divides the sources of evidence into fieldwork, communicative behavior, behavioral effects of processing, physiological effects of processing, and corpus linguistics methods. For present purposes, however, we will be better served by a different classification, one focused on the relation of the evidence to language itself rather than the methods used to gather that evidence. Accordingly, I'll classify sources of evidence into usage data, processing data, and metalinguistic data. *Usage data* consists in linguistic tokens of the phenomenon of interest itself, whether gathered observationally or elicited experimentally. *Processing data* involves both behavioral and physiological markers of linguistic cognition, excluding the bare form of the linguistic tokens themselves. Finally, *metalinguistic data* consists not in subjects' linguistic behavior as such, but in their attitudes and behavior relating to facts *about* language. These last data, of course, include linguistic intuitions, and we'll examine their evidential value at length in what's to come. Here, however, we'll look at some examples of the other two sources of evidence to serve as a basis for comparison.

Usage data is relevant to all topics in linguistics, and can be gathered experimentally or observationally. Linguists doing fieldwork have developed a number of elicitation techniques to gather usage data, from asking leading questions and requesting descriptions of visual scenes, to more complicated techniques such as deliberately producing potentially ungrammatical utterances in order to elicit a correction, or asking consultants fill in a missing word from a sentence in order to see which lexical items can play certain semantic or syntactic roles. These elicitation techniques are the standard methodology for linguists studying languages not spoken by many academics, such as the languages of small societies or informal dialects of minority social groups. Elicitation is much less frequently used to study the syntax and semantics of languages commonly spoken by linguists themselves, such as English, German, and Mandarin, suggesting a general preference for metalinguistic data over usage data[18].

Linguists also gather usage data in bulk to create a corpus of written or spoken language samples. Such corpora can be gathered to target language use in a specific context, as in the CHILDES corpus of spontaneous child-directed speech, or can just gather tokens of language use regardless of genre and context, as with the Google Ngram Viewer collection of digitized books. Most corpora consist not only of the raw data, but also metadata called annotations, added to make possible a wider variety of linguistic analysis. For example, a corpus might tag sentences with basic syntactic structure, in which case it is a treebank (e.g. the Penn Treebank). Some corpora, such as the Philadelphia Neighborhood Corpus, include detailed socioeconomic data about the speakers in the corpus, thus allowing systematic sociolinguistic analysis. Some of the most important research corpora, such as the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC) aim to be useful to a variety of linguistic sub-disciplines, and thus include annotations such as part-of-speech tagging, lemmatization[19], date,

[18] Thus, we use a different methodology to study academic languages than we do to study non-academic languages, with the vast majority of languages falling into the latter camp. This methodological schism begs for justification. If, as I am arguing, metalinguistic judgment data is inferior to usage data, then no tenable justification is forthcoming. So one ramification of my argument is that we should partially close the gap between how we study academic and non-academic languages.

[19] A lemmatized corpus groups words which belong to the same dictionary entry together despite differences or apparent identity in surface form. In a properly lemmatized corpus, for instance, one can

and genre (academic, spoken, fiction, etc.). In fact, given effective text analytic software to perform language identification, part-of-speech tagging, and lemmatization on the fly, any body of digitized text, such as records of social media, can be treated as an annotated corpus.

Corpus analysis is used as a source of evidence in both syntax and semantics. We can study the grammaticality of a particular construction, for instance, by looking at its frequency in a corpus. Consider the split infinitive, i.e. ‘to [adverb] [verb]’ as opposed to ‘to [verb] [adverb]’ (e.g. ‘to boldly go’ instead of ‘to go boldly’). If we can detach ourselves from the trauma of high school grammar class, we probably intuit that the split infinitive is acceptable in Standard American English, an intuition backed up by the construction’s frequent appearance in COCA (Davies 2014). The corpus analysis can take us further, however, documenting the meteoric rise[20] in the use of the split infinitive between the years 1990 and 2012, during which it went from comprising 37.9% of infinitive-adverb constructions to comprising 55.7% (Davies 2014). Thus, the syntactician can learn from a corpus analysis not only that the split infinitive is grammatical, but also that its acceptability appears to be on the rise, despite those traumatic high school grammar classes. They can also assert these conclusions with greater confidence than they could if they had only subjective judgment to rely on, backed as they are by the weight of a 450-million-word corpus who sheer size makes statistical significance trivial.

Processing data differs from usage data in that rather than consisting in the bare form of language in use, it includes behavioral and physiological markers concomitant to linguistic

search for the headword ‘be’ and retrieve instances of ‘is’, was’, ‘were’, etc. Likewise, one could distinguish instances of ‘spit’ (landform) from ‘spit’ (saliva) despite identical surface form.

[20] It struck me as I wrote this that ‘meteoric rise’ is an odd phrase, given that meteors only fall. So I consulted COCA’s sister corpus, the Corpus of Historical American English (COHA), and found, sure enough, that ‘meteoric rise’ has undergone a meteoric rise of its own, from appearing 0.08 times per million words in the 1950’s to its present frequency of 0.47 times per million words. This increase does not track the general use of the word ‘meteor’ which has remained relatively constant over the same time period. A potential explanation for the semantically odd phrase is suggested by the earliest token present in the corpus, an 1894 discussion of the “meteoric rise and fall” of a British aristocrat. “Meteoric” apparently originally evoked brevity, not a path of motion. We might surmise that the phrase became idiomatic, then abbreviated, losing the ‘and fall’, thus yielding the modern semantically opaque idiom.

production and comprehension. These markers include among other things, reaction times, eye movements, and neuronal activity. Consider how Trueswell and Kim (1998) use a self-paced reading measure to determine the factors involved in parsing garden path sentences. A garden path sentence is difficult to process because the parser's expectation for how a sentence will end is violated. Contrast the following sentences (used in Trueswell and Kim):

1. The photographer accepted the money.

2. The photographer accepted that the fire could not have been prevented.

3. The photographer accepted the fire could not have been prevented.

Sentence 3 is a garden path sentence because without the complementizer 'that', the reader is expecting a direct object, not a sentence complement, after "accepted the…" This makes sentence 3 more difficult to process than sentences 1 and 2. The question is whether this effect is due solely to a constraint triggered by syntactic information in the lexical entry for 'accepted'. Trueswell and Kim show that it is not by priming readers with other verbs, some of which, like 'accepted', prefer direct objects, and some of which prefer sentence complements. The preference of each verb was determined by statistics drawn from the Penn Treebank corpus. Trueswell and Kim found that when primed with a verb preferring a sentence complement, reading times on sentences like 3 significantly improved, indicating that the brain's syntax apparatus is rather flexible when reconstructing syntactic structure from surface form. This illustrates one way in which behavioral measures can serve as evidence for claims in syntax[21].

They also serve as useful evidence for semantics and pragmatics. Researchers in the Gricean tradition, for instance, sometimes try to identify which implicatures are conventionalized

[21] Here, briefly, is another. Many leading research programs in syntax posit an underlying syntactic structure different from the surface form actually uttered. Processing data can help us determine whether these proposed underlying structures have psychological reality. Psycholinguists, for instance, can test whether two sentences sharing an underlying feature but not a surface feature prime each other. See, for instance Bock et al. (1992)

and which involve inference on the part of interlocutors. On the assumption that on-the-fly pragmatic inference requires more processing than merely accessing a linguistic convention, we can use processing data as evidence for whether an implicatures is conventionalized. Atanassov et al. (2013), for instance, consider the scalar implicature 'might' implies 'not must'.[22] They use eye-tracking visual world experimental paradigm, where researchers can tell which interpretations a subject is considering by monitoring gaze. In Atanassov et al.'s study, they found a significant processing delay when stimuli contained a scalar 'might' implicature when contrasted with the corresponding implicature-less 'must' sentences. This result, they argue, is evidence that hearers engage in real time pragmatic inference with scalar implicatures, and so these implicatures are not conventionalized in the same way that lexical meaning is.

In addition to reaction times and eye movements, processing data includes other classic psycholinguistic measures such as self-paced reading behavior, and neurolinguistics measures gathered using electrodynamic and hemodynamic techniques. Between these diverse sorts of processing data and the various means of obtaining usage data, linguistics has no shortage of relevant data to use as evidence.

Intuitions about acceptability and truth

The third type of evidence available to linguists is metalinguistic data, which comprises speakers' introspective, reflective, and behavioral access to facts about language. These can be facts of almost any relevant sort: about acceptability, degrees of grammaticality, truth, entailment, ambiguity, what is said rather than implied, semantic anomaly, perceptual distinctiveness, word and morpheme boundaries, adequate paraphrases and translation, etc. Note that these are not necessarily metalinguistic facts; what makes this data metalinguistic is that subjects engage in a metalinguistic task (paraphrase, judgment of acceptability, etc.) to produce it. Some varieties of metalinguistic data are uncontroversial. Researchers in the field can't avoid having to ask native

[22] On a Gricean account, this implicature arises because 'must' is more informative than 'might', so it would be uncooperative of a speaker to use 'might' if they believed 'must' to be true.

speakers of the languages they document to give them paraphrases and translations, for instance. Consequently, we won't be putting metalinguistic data under scrutiny as a class. Instead we'll focus on the types of metalinguistic data most important to syntax and semantics: acceptability judgments and truth-conditional judgments[23]. From here on out, discussion of intuitions will, unless otherwise noted, refer only to acceptability judgments and truth-conditional judgments.

The word 'intuition' is controversial when applied to these phenomena, with some authors preferring 'judgment' or even, in one case, 'knowledge' (Ludlow 2011b). I will use 'intuition' and 'judgment' interchangeably. The referent in any case is clear enough. Linguistic intuitions are the consciously-accessible but reflexively-produced metalinguistic assessments of a competent language-user.

The metalinguistic assessments most central to contemporary linguistics are *acceptability judgments*. Syntacticians are interested, among other things, in determining the rules governing well-formedness in natural languages, and intuitions about whether a particular sentence is acceptable are meant to speak to whether that sentence is well-formed. For example, a competent speaker of English will report that 1 but not 2 is grammatical.

1. The cat is on the mat.
2. *The cat the mat on is.[24]

This acceptability judgment serves as evidence that 2 violates the rules of the grammar of English, thus providing the syntactician with a datum: the syntactic rules of English permit 1 but not 2. The linguist can then use this datum to support a particular account of what the syntactic

[23] Syntax and semantics make use of other types of metalinguistic data as well, such as judgments about ambiguity and judgments about what is said. I think most of the discussion in this chapter applies to them as well, but I will not be addressing them specifically.

[24] Following the convention in linguistics, the asterisk denotes an ungrammatical, or at least unacceptable, sentence.

rules of English actually are.

To illustrate, consider a more serious example, from Chomsky's *The Logical Structure of Linguistic Theory*, discussed in Wasow and Arnold (2005). Chomsky (1975, 474-477) appeals to the intuition that 3a and 3b are grammatical, but 4 is not.

3a. The detective brought in the suspect.

3b. The detective brought the suspect in.

4. *The detective brought the man who was accused of having stolen the automobile in.

At question is which transformations involving moving a preposition (as in the difference between 3a and 3b) are allowed in English. Chomsky appeals to the intuition that 4 is ungrammatical to give evidence for a principle that "the separability of the preposition is determined by the complexity of the NP object." Roughly, because "the man who was accused of having stolen the automobile" contains an embedded relative clause, it is too complex to allow the transformation.

I don't share Chomsky's intuition that 4 is unacceptable, but we need not take the asterisk to denote absolute unacceptability. Most syntacticians believe that acceptability comes in degrees (and in fact there are conventional symbols such as ‘??’ which denote lesser degrees of unacceptability), and frequently the asterisk indicates not merely the unacceptability of an isolated sentence, but the unacceptability of a sentence relative to a contrast sentence. So even if we think 4 is clearly acceptable, we might construe Chomsky as identifying a strong intuition that 4 is less acceptable than 3b. A fact about relative acceptability, of course, is a different sort of evidence than a fact about bare unacceptability, since they have different theoretical import. For sake of exposition, then, let’s grant to Chomsky the bare unacceptability of 4, but it’s worth keeping in mind that acceptability judgements are about relative acceptability at least as often as they are about bare acceptability.

Wasow and Arnold observe that Chomsky’s aforementioned argument includes not only acceptability judgments, but a second type of syntactic intuition, which they call *secondary intuition*. Where primary intuitions are binary judgments about the (relative) acceptability of a sentence, secondary intuitions are judgments about why a sentence is or isn't acceptable. In the above example, Chomsky not only appeals to an intuition *that* 4 is unacceptable, and so likely ungrammatical, but an intuition *why* it is: namely, that it is too complex. It isn't clear what role Chomsky's appeal to secondary intuition is playing, however. It may be that he intends it to serve as further evidence for his theory, or merely that he uses it as a means of formulating a hypothesis, and only the acceptability judgment is meant to serve a justificatory role. Either way, what role primary and secondary syntactic intuitions can legitimately play in linguistic theory is an important question. Since secondary intuitions are likely even less reliable than primary intuitions, however, I will focus on the latter.

Not all linguistic intuitions are syntactic, of course. The syntactocentrism of generative linguistics has meant that both the use of intuitions in linguistics and the metatheoretical discussion of their role has focused on syntactic intuitions, but intuitions are central to the practice of other domains as well. In particular, some branches of semantics rely on intuitions as a primary source of evidence much in the same way that generative syntax does. Intuitions in truth-conditional semantics are not about the well-formedness of sentences, but about their truth conditions. In practice, determining truth conditions often involves assessing intuitively the truth of a sentence in a given scenario, yielding a *truth-conditional judgment*. To illustrate, consider the following argument that the English 'if...then...' conditional cannot be equivalent to the material conditional:

Lewis Carroll (1894) presents a scenario where three brothers, Allen, Brown, and Carr, jointly run a barbershop according to the rule that at all times at least one of the three will occupy the shop. Consider whether 5 is necessarily true given this scenario:

5. If Allen is out, Brown is in.

Intuitively, 5 is false given the scenario. This truth-conditional judgment is the sort of semantic intuition which plays an evidentiary role in formal semantics. We can also explain why 5 is false—it is possible for both Allen and Brown to be out, so long as Carr is in the shop. Perhaps this explanation comes from some sort of secondary semantic intuition, analogous to the secondary syntactic intuitions we identified above, but more likely the explanation is primarily the result of non-intuitive reasoning processes. Either way, the primary semantic intuition plays the more important role as evidence in this argument. Our intuition is that 5 is false in the given scenario, which means that its negation must be true.

6. It is not the case that if Allen is out, Brown is in.

Given our intuition that 5 is false and the assumption that "It is not the case" functions as logical negation, 6 must be true. Now suppose that English "if...then..." functions as the material conditional. Let a = Allen is in, b = Brown is in, and c = Carr is in. We can then represent 6 in the propositional calculus as in 6p.

6p. $\sim (\sim a \rightarrow b)$

We can rewrite the conditional as a disjunction, yielding 6p'.

6p'. $\sim (a \,||\, b)$

By DeMorgan's law, we get 6p''.

6p''. $\sim a \,\&\, \sim b$

So if 6 is true, then ~a. In other words, if 5 is false and English "if...then..." corresponds to material implication, then it must be the case that Allen is out. This entailment is false, since it is consistent with the stipulated scenario that Allen is in. So either 5 isn't false, or "if...then..." isn't the material conditional. Our intuition is evidence that 5 is false, so it must be the case that "if...then..." isn't the material conditional.

Note the key role our intuition played in the argument. Intuitions about whether a sentence is true in a given scenario frequently play similar roles in formal semantic argumentation. Truth-conditional judgments are not, of course, the only semantic intuitions—we have intuitions about the meanings of individual lexical items, for instance—but they will be our focus given their particular importance to the science of semantics.

The evidential role of linguistic intuitions

I have presented acceptability and truth-conditional judgments as functioning in an evidential role in syntax and semantics, but this is a claim in need of some defense. In fact, I will ultimately reject it, at least in part. It will be helpful, however, to try to understand why we would think that these intuitions are scientific evidence.

Mental events of the sort which might reasonably be called intuitions are put to a variety of uses in the sciences, and most of these uses are less controversial than the use of intuitions as evidence. For example, hunches and intuition-guided guesses can help a scientist formulate a hypothesis or decide which research path looks to be more fruitful. In the early stages of developing a new theory, a scientist may use her intuitions to temporarily fill theoretical gaps. Intuitions can also allow the scientist to make fast, non-transparent jumps in reasoning because of years of experience with the subject matter. Appeals to intuition might sometimes play important rhetorical and pedagogical roles in scientific discourse. All of these uses of intuitions are valuable to scientific practice. Linguistic intuitions plausibly play most of these roles for linguists. At question is whether they also function as evidence used to confirm and disconfirm linguistic theories.

As we saw in the previous section, acceptability and truth-conditional judgments *can* be presented as evidentiary support for theoretical claims, and prominent linguists sometimes claim that intuitions do in fact constitute confirmatory evidence. "In actual practice," Chomsky writes, "linguistics as a discipline is characterized by attention to certain kinds of evidence that are, for the moment, readily accessible and informative: largely, the judgments of native speakers" (1986, 36). This is a commonly held attitude in syntax. A typical argument for a claim about syntax, such as the example in the previous section, includes an appeal to acceptability judgments about sample sentences. In an especially hyperbolic mood, syntacticians have even claimed that when it comes to grammaticality, "[a]ll the linguist has to go by...is the native speaker's intuitions about language" (Haegeman 8). Claims of this sort are exaggerated, since it is clear that other sorts of data are accepted and commonly used in generative linguistics. But even many linguists who acknowledge the existence of other sorts or relevant evidence see intuitions as central to linguistic confirmation. Psycholinguists Lila and Henry Gleitman, for instance, argue that "[t]he mental events that yield judgments are as relevant to the psychology of language, perhaps, as speech events themselves" (1979, 105). The attitude that linguistic intuitions are an important sort of evidence for the science of language is widely shared and informs scientific practice.

This brings us back to our motivating puzzle. Intuitions are not generally counted as empirical evidence in modern science, so inasmuch as linguistics does rely on intuitions as evidence, it seems out of place in the constellation of sciences. That this is a real concern is demonstrated by the fact that throughout the history of linguistic science, critics from within linguistics have expressed reservations about the practice of using intuitions as evidence (e.g. Bloomfield 1933, Harris 1954, Labov 1996, Wasow and Arnold 2005). These reservations require a response, and the aim of the response should be to demonstrate that intuitions about language, in contrast to intuitions about, say, chemistry or economics, have a distinctive feature or features which enable them to function as scientific evidence.

We'll look at three such features which linguists and philosophers have suggested distinguish intuitions about language: (1) That intuitions about language, as opposed to non-scientific intuitions elsewhere, *lead to fruitful scientific discourse*. (2) Linguistic intuitions *have a close causal relationship with language*, whereas intuitions in other disciplines do not typically have a close causal relationship with their subject matter. (3) Intuitions in linguistics are *reliable*, albeit for somewhat mysterious reasons, but intuitions in other disciplines are typically unreliable. Each of these distinctive features—fruitfulness, apt etiology, and reliability—would explain why intuitions in linguistics could have scientific status as evidence, and thus would allow us to reject (1) from our original trilemma. We'll scrutinize each in turn.

Fruitfulness

One means of justifying the scientific status of linguistic intuitions is to appeal to their role in producing fruitful scientific research programs. Gross and Culbertson (2011, 654) provide a recent example of using this justification rather than an etiological one, acknowledging that "the relation between linguistic competence and the cognitive capacities recruited in meta-linguistic judgments remains obscure." They argue instead that one main justification for using intuitions as evidence comes from "the continued success and fruitfulness of the theories based upon them." The argument, I take it, is straightforward: if a scientific methodology leads to fruitful theories it is justified, syntax has produced fruitful theories, and syntax relies on intuition-evidence, therefore intuition-evidence is justified.

I don't want to deny that syntax has produced fruitful theories. I reject, however, the premise that if a scientific methodology leads to fruitful theories it is justified. I take fruitfulness to a weak, defeasible consideration in methodological justification, with more rigorously epistemic considerations like accuracy and consistency having priority. Agreement or disagreement with evidence known to be reliable will usually trump appeals to fruitfulness when it comes to deciding which sorts of evidence to use. In large part, this is because fruitfulness suggests but hardly entails reliance on good evidence. Consider your least favorite major religion. It's a safe bet that theologians of that religion have been having fruitful inquiry for centuries. They publish prodigiously, engage in persuasive argumentation, look back on the history of their discipline and feel like they have made progress, and have reached consensus on a number of the most central issues in theology. From the theologian's perspective, their field constantly appears to "disclose new phenomena or previously unnoted relationships among those already known," to use Kuhn's definition of scientific fruitfulness (1977, 321). All this fruitfulness, and their sources of evidence aren't even evidence for anything real! The fruitfulness of their theological science is hardly a convincing reason to take their scripture or revelation as good scientific evidence. Fruitfulness is thus a scientific virtue of lesser importance, more telling in its absence than its presence. A lack of fruitfulness might be damning for a particular scientific methodology, but its presence guarantees little. That generative syntax and formal semantics have been fruitful, at least to some extent, is thus not a particularly good reason to take the use of intuition-based evidence to be a successful methodology[25]. We'll turn our attention to more plausible defenses of linguistic intuitions.

Apt etiology and the ontology of language

Another means of demonstrating the empirical status of linguistic intuitions is to show that they have a relevant empirical etiology. This would require showing that the intuitions themselves have a close causal relationship with language itself, since language is the object of scientific interest. Linguistic intuitions, the argument runs, are fairly direct causal products of language itself, so they carry empirical information about language. Most other intuitions about scientific objects, however, do not have an origin in the objects themselves; intuitions about electrons are not principally the product of electrons, so we can't study subatomic particles by using intuition as evidence. Linguistics' anomalous use of intuition-evidence is thus explained by the unique way in which linguistic intuitions are related to language. As I discussed in Chapter 1, however, there are multiple valid conceptions of language, and this means that there can be multiple etiological defenses of language.

In particular, there will be etiological defenses which attempt to attribute the etiology of

[25] The argument from fruitfulness also loses its fangs if it turns out that linguistic intuitions don't actually play as central an evidential role in linguistics as is sometimes claimed. I will argue that this is the case towards the end of this chapter, but that argument will take as a starting point that I have demonstrated that intuitions are not good scientific evidence. While I don't think that appealing to the conclusion of that argument (that intuitions are not typically used as evidence in practice) here would be viciously circular, it might appear to some readers to be so, so I avoid resting my case on this alternative argument against the fruitfulness approach.

intuitions to language the external social object, and defenses which attribute their etiology to something like the language faculty—language the psychological object. I'll address both, in turn. I won't discuss defenses of intuitions proposing an etiology of language as some sort of mathematical object. As far as I can tell, no one in the present literature on linguistic intuitions takes that line, and even if there were, say, a Platonist about language who did, their account of linguistic intuitions would be clearly a priori, and thus not resolve our motivating trilemma anyway.

The etiological defense—social

Most contemporary defenders of the evidential status of linguistic intuitions take the primary subject matter of linguistics to be a psychological entity, and their argument from etiology accordingly focuses on the causal connection between intuitions and the psychological faculty of language. This is the case largely because linguistic intuitions are primarily used as evidence in generative linguistics, and generative linguists are typically committed to a psychological ontology of language. However, at least one notable defender of linguistic intuitions, Michael Devitt, has broadly anti-generativist commitments Accordingly, he attempts to defend intuitions on the basis of their causal connection to the social object that is language. The social etiological defense is an approach worth considering, so despite the relative unpopularity of Devitt's account, I'll take some time to address it.

Devitt's basic causal story appeals to the expertise held by all of us, who live in a language-saturated world. On his picture (2006b, 497-98), a competent speaker "is surrounded by tokens that may...be grammatical, be ambiguous, corefer with a specific noun phrase and so on." Experience with these tokens gives her the material to form a sort of folk theory of language if she is reflective about her linguistic experiences. Internalizing this folk theory allows her to "judge in a fairly immediate and unreflective way that a token is grammatical, is ambiguous, does corefer." Linguistic intuitions, as Devitt sees it, are nothing more than these immediate and unreflective judgments. Devitt (2006b, 499) allows that the language faculty may play some factor in shaping an individual's theory of language, but only in a highly mediated, indirect way. The causal relationship between the social facts of language and a particular linguistic judgment is more direct: speakers synthesize as a personal theory the linguistic facts in their environment, then reflexively access that theory in the form of an intuition.

As mentioned, driving Devitt to this account of linguistic intuitions is an anti-psychological account of the ontology of language (Devitt and Sterelny 1987; Devitt 2006a). His etiology for intuitions fits neatly with social ontologies of language since according to Devitt experience with the external signs of language is what seeds linguistic theories. This does not mean that his account is incompatible with psychological accounts of the ontology of language, however. If we take linguists to be primarily interested in I-language, we can still accept this as a good scientific justification of linguistic intuitions. Intuitions would be at best only indirect evidence for I-language, but nothing in psychological linguistic ontologies commits us to denying this. Devitt's etiological defense is thus compatible with either ontology of language, even if it fits better with a social ontology. Therefore, the validity of Devitt's account doesn't hang on the question of the metaphysics of language.

It hangs instead on two issues. First, supposing Devitt has accurately identified the etiology of linguistic intuitions, does this etiology justify their use as scientific evidence? Second, and more fundamentally, is Devitt's causal story supported by the evidence? Sorting through these two questions isn't straightforward, because there's more than one way to fill out the details of Devitt's account. To fill it out, let's gather more clues from what Devitt writes.

Surprisingly, perhaps, given his fundamental disagreements with generative linguists, Devitt (2006b) is keen to give justification of their standard methodology. His starts with the fact that in actual practice syntacticians rarely call upon the intuitions of the folk. While folk intuitions are sometimes gathered, and information sometimes drawn from corpora or experiments, the bulk of the data come from the intuitions of linguists themselves. It's a reasonably safe bet that the 'informants' referred to in any syntax paper include the author's colleagues and likely the author herself. This practice of drawing predominantly on professional linguists' intuitions has come in for some criticism (e.g. Dahl 1979; Wasow and Arnold 2005; Gibson and Fedorenko 2010), but Devitt thinks not only that the practice is unproblematic, but also that it grounds the epistemic worth of intuitions in linguistics. After all, intuitions are the unreflective applications of theories of language, and don't linguists have the best theories?

As Devitt sees it, even the folk have intuitions that are likely to be true, but their intuitions are outclassed by those of linguists. Consider the source of linguistic intuitions in Devitt's picture: the individual observes her own utterances and those of her interlocutors and through reflection and generalization comes up with a linguistic theory. Given that the data she generalizes from is reliable and relevant, the theory is likely to be reasonably accurate. And since her intuitions are applications of this theory, they too are likely to be reasonable accurate. Folk theories, however, are not going to be nearly as accurate as theories tempered by the scientific process, so "the intuitions that linguistics should most rely on are those of the linguists themselves, because linguists are the most expert" (2006b, 499). If this argument is sound, not only are linguists justified in using intuitions as evidence, but they also already tend to make use of intuition-based evidence in the best possible way.

Devitt's argument isn't sound. To see this, we'll consider three possible causal stories relating personal theories of language, intuitions, and evidence, and see that none of them ground an effective Devitt-esque apology for the evidentiary use of intuitions. It's not entirely clear which of the three stories Devitt would accept, but it doesn't matter since all fail.

Suppose that the relationship between learned theories and intuitions is straightforward: the speaker possesses a theory of grammatical correctness, unconsciously applies that theory to a sample sentence, and unconsciously produces a judgment of whether that sentence is grammatical according to the theory. Should the speaker's judgment in this case count as scientific evidence for a syntactic theory? Clearly not. In this case intuitions are merely going to confirm the theory the speaker already possesses. If the speaker is a non-linguist, this means that the intuitions are at best evidence we can use to reconstruct the speaker's folk theory of language. To take the layperson's intuitions as evidence for the structure of language or the language faculty would thus be to put more trust in folk linguistics than is warranted.

What if the intuitions are those of a trained linguist? If anything the situation is worse, since if the linguist already accepts the theory she is trying to provide evidence for, the theory itself will shape her intuitions. Taking her intuitions as evidence will thus lead to an especially pure kind of confirmation bias, in which the theory under consideration causes her intuitions, which are in turn used to support the same theory. I'm far from the first to note this problem. Dahl writes, "It is well-known among linguists that intuitions about the acceptability of utterances tend to be vague and inconsistent, depending on what you had for breakfast and which judgment would best suit your own pet theory" (1979, quoted in Schütze 1996, 113). So *pace* Devitt, in cases where a straightforward social etiology of linguistic judgments is true, intuitions—especially those of linguists—should not be used as evidence to confirm or disconfirm theories.

In at least some cases the effect of learned theory of language on metalinguistic judgments is more indirect. Dabrowska (2010), for example, shows that generative syntacticians are more likely than others to intuit that sentences containing long-distance dependencies are acceptable, despite the fact that their theory denies that they are grammatical. She postulates that this is due to generative syntacticians' frequent exposure to sentences containing long-distance dependencies in the articles they read, which inoculates them to the sentences' ungrammaticality. In this case the relationship between learned theory and metalinguistic judgment is indirect; the judgments are not merely straightforward applications of the theory. Devitt's argument fares no better in these cases, however. Devitt's justification of intuition-based evidence requires a direct connection between learned theory of language and intuitions. Linguists' intuitions are reliable, he argues, because linguists have good theories. If those intuitions aren't straightforwardly in accord with those theories, then the goodness of the theories is irrelevant. In cases where the effect of learned theory of language on intuitions is indirect, then, Devitt's argument fails to justify the appeal to intuitions as evidence.

I have presented a dilemma for Devitt's case. If intuitions are direct applications of learned theories, then using them as evidence would be to embrace confirmation bias. If they are not applications of learned theories, then we can't appeal to the reliability of the theories to explain the reliability of the intuitions. A possible way out of the dilemma is to argue that intuitions are not direct applications of the theory, but the output of unconscious deductions from the theory. Consider an analogy with mathematical reasoning. The mathematician is aware of and has internalized a set of axioms and theorems. In a reflexive, unconscious process, her brain deduces that another theorem, θ, logically follows from those she already knows. The output of this process is a hunch that θ, and the mathematician conjectures that θ. It seems implausible that linguistic theories have deductive properties of the same sort as mathematical theories, but for sake of argument let's suppose they do. Linguists have the best theories, and their intuitions may be reflexive, unconscious deductions from already accepted theorems to new implications of those best theories. This account avoids both horns of the dilemma, since it involves no confirmation bias, and since the truth of a logically implied theorem depends on the truth of the theory implying it.

Even supposing this account accurately described any actual metalinguistic intuitions, we wouldn't accept those intuitions as evidence. Imagine what would happen if our fictional mathematician tried to publish a paper arguing for θ, using the fact that she had a hunch that θ as her primary evidence. In mathematics hunches play an important role in guiding inquiry, but the proof is in, well...the proof. If linguistic intuitions were analogous to mathematical hunches the same standards of evidence would apply. Intuitions would be used to create hypotheses, but if the intuited facts were actual deductions from accepted theory, the linguist would be expected to demonstrate the deduction. So even though appealing to unconscious deduction allows us to avoid the dilemma for Devitt's argument, it doesn't justify the use of intuitions as evidence.

Perhaps neither of these interpretations is charitable to Devitt. It may be that he sees the etiology of intuitions not as application of an internalized theory in any of the ways we've just canvassed. Instead, he might see subjects not so much as theorists, and more as recording devices. A language user soaks up, like a sponge, facts about how language is used around him. In fact, he soaks up social facts of all sorts: how people dress, what music the in-crowd listens too, the means different people use to avoid colliding on the sidewalk. When we elicit his intuitions about a social fact, including a fact about language, we aren't so much accessing his worked-out theories as we are just giving the sponge a squeeze and letting the absorbed data drip out. Humans, as social animals, reliably track the social facts around them, and linguistic intuitions access the language-specific subset of these facts.

Unfortunately, the human brain, although rather spongey in a literal sense, doesn't metaphorically soak up the social facts well enough—it is not a reliable enough recorder of social facts to ground this account. If it were, we could do not only linguistics but all sorts of social science from the armchair. But we can't. Sociologists have to gather data and use sophisticated statistical techniques precisely because the human brain's passive data gathering is not extensive enough, reliable enough, or statistically sophisticated enough to serve as the basis for scientific theorizing about social facts.

This is demonstrably true when it comes to sociolinguistic facts. Consider an illustrative case from the early history of modern sociolinguistics. Labov (1996) recounts how he became aware of the disconnect of metalinguistic reports from actual usage when a phone survey intended to gather data on regional dialects yielded results known to be false. "The erratic scattering of responses obtained through introspection," he discovered "has little relation to the clear regional patterns produced by mapping behavior" (1996). Spurred by this failure of introspective reports to correspond with linguistic performance he developed field methods to gather the same data, and these field methods confirmed that first-person metalinguistic reports are no substitute for usage data. A telling example is case of positive 'anymore' in Philadelphia. In most dialects of English, 'anymore' is a negative polarity item, but in some East Coast dialects it can appear in sentences with positive polarity with a meaning of roughly 'these days' or 'nowadays'. While interviewing subjects in their homes, Labov found that many would deny using positive 'anymore', knowing what it means, or even having ever heard it used[26]. In these same interviews or in follow-ups the exact same subjects often made use of positive anymore:

> Jack Greenberg, a 58-year-old builder raised in West Philadelphia, gave introspective reactions that were so convincing that I felt that I had to accept them as valid descriptions of his grammar. Yet two weeks later, he was overheard to say to a plumber, "Do you know what's a lousy show anymore? Johnny Carson." A 42-year-old Irish woman said, "I've never heard the expression." Earlier in the interview she had said, "Anymore, I hate to go in town anymore," and a short time later, "Well, anymore, I don't think there is any proper way 'cause there's so many dialects.

These subjects made valid use of positive 'anymore', so they must have had tacit knowledge of its semantic and syntactic properties. Their metalinguistic judgments, however, did not make use of this knowledge. At the very least, examples like this suggest that subjects, even to the extent that they are reliably tracking their sociolinguistic environment, do not have reliable intuitive access to that data.

Summing up, no matter how we cash out a social etiology for linguistic intuitions, they won’t serve as the type of reliable data we would accept as scientific evidence. Moreover, it’s not clear that linguistic judgments really are predominantly produced by reliably observed social facts. Devitt’s etiological defense, while intriguing, is thus on poor footing. We turn, then, to the other major contender.

The etiological defense—psychological

A more widely accepted account of the etiology of linguistic intuitions attributes their

[26] Labov found no stigma attached to the use of positive anymore, so it is unlikely that these responses were evasive or dishonest.

production in large parts to the psychological faculty of language, or the linguistic competence. Devitt (2006a; 2006b) dubs this account "Voice of the Competence," and we will follow his nomenclature.

> VOICE OF THE COMPETENCE (VoC): Metalinguistic judgments are principally the product of the mental mechanisms responsible for accurate linguistic performance.

VoC has its origins in Chomsky, especially his (1986), but philosophers who have articulated versions of the argument more recently include Fitzgerald (2010) and Maynes (2012), among others. On their account, linguists are interested in studying a particular cognitive object, the language faculty or linguistic competence. Linguistic intuitions emerge from a process involving the competence: The competent speaker considers a sentence, which serves as input to the language faculty. The language faculty then processes the sentence and in a fast, unconscious process determines whether the sentence is grammatical (or the sentence's truth-conditions, etc.). This determination is signaled to conscious cognitive systems, yielding the intuition about the sentence. Given the close causal relationship between the intuitions and the workings of the language faculty, the intuitions must carry information about linguistic competence. Intuitions thus provide evidence for the structure of linguistic competence, which is the target object of study

Proponents of VoC acknowledge that the information carried by intuition is noisy. In *Knowledge of Language*, Chomsky points out that "[i]n general, informant judgments do not reflect the structure of the language directly; judgments of acceptability, for example, may fail to provide direct evidence as to grammatical status because of the intrusion of numerous other factors" (1986, 36). These other factors include judging a sentence to be unacceptable for semantic or sociolinguistic reasons rather than a lack of syntactic well-formedness, as well as rejecting a sentence because it is too complex to parse easily even though it is in principle parseable. Although this noise interferes with the quality of the evidence, noisy evidence is the norm in science, not an insurmountable problem. Moreover, usage data in linguistics is noisy in the same way. Linguistic performance originates in the language faculty, but also involves the intrusion of other factors, including those selfsame semantic and social factors as well as limitations on memory, motor control, etc. This very fact is what motivated the competence/performance distinction in the first place. The presence of noise alone thus fails to put metalinguistic data on weaker ground than usage data, when it comes to studying the language faculty. Noise or no noise, according to VoC the primary origin of linguistic intuitions is linguistic competence, so intuitions carry valuable information.

If true, VoC gives compelling reason to accept the reliance on intuitions as evidence in linguistics. We have good reason, however, to question the claim that the language faculty is the primary origin of linguistic intuitions. Defenders of VoC, when they offer any argument in favor of the view, typically appeal to an inference to the best explanation (e.g. Maynes and Gross 2013). Doesn't it make sense that the etiology of linguistic judgments is to be found in linguistic competence? But this inference holds little water, since it is unclear why the language faculty would be in the business of issuing metalinguistic judgments. The job of linguistic competence is to produce and interpret utterances, and making metalinguistic details about these utterances consciously accessible is generally neither necessary nor helpful in accomplishing this task. Both defenders and critics of VoC acknowledge this. Maynes and Gross, who accept VoC, admit that "the capacity for linguistic intuitions is a further, indeed dissociable, capacity that goes beyond the capacity for language production and comprehension" (2013, 717). Devitt, who rejects VoC, agrees, noting that "the data provided by competence are *linguistic expressions* (and the experiences of using them) not any *observational reports* about those expressions" (2010, 836). In other words, if VoC were true, it would be a substantive and even somewhat surprising fact—the kind of fact we require good empirical evidence to substantiate.

As far as I am aware, good empirical evidence of the causal connection between linguistic competence and linguistic intuitions has yet to be produced. Even commentators sympathetic to VoC sometimes admit this. Fitzgerald[27], for instance, writes that "[w]e don’t know how conscious judgements are derived, or the mechanics of the role the linguistic systems play in issuing in these judgements" (2010, 144). And Culbertson and Gross acknowledge that "little is currently known about the causal etiology of linguistic intuitions" (2013, 720). These are recent, scientifically-informed works defending the use of intuitions as evidence. Their acknowledgement that we don’t have much direct evidence supporting their etiological story is thus a believable assessment of the state of the field.

Let’s take score. The highest number of points in favor of VoC would come from direct empirical evidence that it accurately describes the causal origins of linguistic judgments, but even its defenders acknowledge that we have almost none. No points for VoC from that domain, then, but no points for any alternative, either. VoC could also have scored by a good argument from function, but it turns out that we have little reason to think that the function of the competence includes issuing metalinguistic data. As far as I can tell, that leaves VoC with only one other chance to score: some sort of indirect evidence, such as an inference to the best explanation. What else could be the principle cause of linguistic intuitions if not linguistic competence?

My answer to this question is that linguistic intuitions could result from what I will call *learned theories of language*.

LEARNED THEORY OF LANGUAGE (LTL): Metalinguistic intuitions are primarily shaped by knowledge of theories of language.

In presenting this alternative, I mean to undermine the inference to the best explanation in favor of VoC. I will, in addition, present empirical evidence in favor of this alternative, thus putting it ahead of VoC on the scoreboard. I use 'theory' loosely, and by "learned theory of language" I mean all the purported facts and generalizations about language an individual learns except

[27] Fitzgerald denies that he (and most linguists) subscribe to VoC, but this denial seems to result from misunderstanding what Devitt means by "Voice of the Competence" (see Devitt 2010). He clearly accepts a position which would fall under the broad characterization of VoC we have given here.

those actually involved in acquiring linguistic competence. Sources of these theories will be diverse. Individuals will pick up and internalize purported facts about language from magazine articles and high school grammar textbooks as readily (and more frequently) than they would from linguistics courses. Much of learned theory of language will derive from the individual's own experiences using language, as well casual discussion about language[28], just as we tend to develop our own personal sociological and psychological theories based on our experiences interacting with other humans. Metalinguistic judgments might then be the result of application of these learned theories of language to linguistic samples.

Studies of folk linguistics have found that folk theory of language often diverges widely from scientific theories of language. Miller and Ginsberg (1995)[29], for instance, collected statements from subjects in the process of formal second language training, since these subjects had ample reason to reflect on the nature of language. They found that subjects' folk linguistic theories demonstrated a number of notable differences from scientific linguistics. For example, subjects tend to equate competence in a language with merely possessing a vocabulary and knowing some syntactic rules. They also tend to resist acknowledging linguistic variation of most sorts, insisting that there is only one right way to say something, usually the way found in a textbook or dictionary. In fact, they would often insist on the dictionary definition of a word even when confronted with native speakers using the word more flexibly. Likewise, they tended to believe that there was only one correct version of a language. Ironically, these second language learners characterized native speakers of non-classroom dialects—in other words, all native speakers—as "bad," "wrong," "incorrect," "guttural," and "inauthentic" (1995: 301). Professional linguists, of course, would disagree with all these folk metalinguistic assessments. Moreover, folk linguistic theories contain very few claims about the sort of things linguists actually try to explain.

[28] This is frequent, and comes in a variety of forms, from speculation about the differences between regional dialects between friends at a bar, to the policing of "bad grammar" on online forums, to discussion between coworkers about whether the phrasing of a memo was "proper English."

[29] See also Niedzielski and Preston (2003).

The folk’s theories are silent about island constraints on wh- movement, on the de re/de dicto distinction, and so forth, but have a lot to say about whose language is better than whose. Folk theory of language, in short, is not the sort of thing that will serve as a good starting point for scientific linguistics.

The danger, then, is that intuitions about language reflect internalized folk theory of language rather than the operations of the language faculty. If this is the case, then linguistic intuitions are as unreliable as the erroneous folk theories tacitly producing them. Note that knowledge[30] of *theories* of language is importantly distinct from knowledge of language. The former is rarely involved in production and comprehension, but the latter always is. Additionally, knowledge of incorrect theories, as we’ve seen, can lead to false beliefs about language, but knowledge of language itself is factive with regards to language. LTL claims that intuitions stem from knowledge of theories of language, and VoC claims that intuitions stem from knowledge of language itself.

At first glance, LTL seems at least as plausible a candidate explanation for the etiology of linguistic intuitions as VoC, thus undermining the inference to the best explanation which might have justified VoC. More importantly, the connection between LTL and metalinguistic judgments is supported by a decent body of empirical evidence, giving it an edge over VoC. The primary sort of evidence in favor of LTL is that subjects' metalinguistic judgments vary in accordance with the sort of learned theories of language they have been exposed to. To see this, we need data contrasting subjects who share competence in the same language, but have different learned theories of language.

The most ready-to-hand case of contrast—literate vs. illiterate adults—provides a striking example of how learning a theory of language (as opposed to merely learning a language) shapes linguistic intuitions. A number of scientists have run experiments with subjects competent in a wide variety of languages and found that literacy, which involves explicitly learning some

[30]'Knowledge' in the linguist's sense (i.e. Chomsky 1986), not the epistemologist's.

metalinguistic theory, significantly shapes linguistic judgments[31]. A series of studies has found, for instance, that illiterate adults segment words into phonemes differently than literate adults. An illiterate English speaker, for instance, might not agree with most literate speakers of English that the word 'cat' is composed of three sound units. This effect has been documented, to give a few examples, among illiterates in Portuguese (Morais et al. 1979), Brazilian Portuguese (Bertelson et al. 1989), Spanish (Adrian et al. 1995), and Serbo-Croatian (Lukatela et al. 1995). In this last study, for instance, less-literate speakers of Serbo-Croatian agreed with the nearly unanimous judgment of literate speakers only 39.3% of the time, which is particularly striking given that there are only a few plausible options in a phoneme counting task. The best explanation for this effect seems to be that learning to read involves acquiring a specific learned theory of language, which in turn drives metalinguistic intuitions.

Not only do illiterates segment words into phonemes differently, but they also differ from literates in how they segment sentences into words. Gombert (1994) found that adult illiterates did not identify the same word boundaries as literates when given sentences to analyze (literates get the right answer 83% of the time; illiterates only 25%), and found some evidence that syntactic rather than phonological factors drove difference. Tellingly, the illiterates could be brought to perform the task identically to literates with task-specific training—explicit teaching of a theory of language nullified the metalinguistic difference. Kolinsky et al. (1987) found a similar effect of literacy on judgments of relative word length among less-literate speakers of French whose performance on the task was at chance level. Again, the theory of language literates pick up while learning to read seems to shape their metalinguistic judgments.

Of course, phonological and lexical identification are not the most important sort of intuition in linguistics. What we really need to undermine VoC is evidence that grammaticality judgments and semantic intuitions are driven by learned theory of language. Fewer studies on

[31] Kurvers et al. (2006) give a good review of this literature, including several of the examples I make use of here.

these types of metalinguistic task have been published, but the ones which have been support LTL rather than VoC. Luria's (1976) famous studies on syllogistic reasoning among illiterates, for example, demonstrate that learning the theory of language involved in literacy changes speakers' intuitions about semantic entailment relationships between sentences. More recently, Kurvers et al. (2006) show among speaker of several North African languages a literacy effect on semantic judgments. For a word-referent discrimination task, literates performed at 43% accuracy, while less-literates were at 17%, and for a syllogistic inference task, literates were at 67% compared to 18.4% for less-literate adults.

As for grammaticality judgments, Karanth and Suchitra (1993) show that among adult speakers of Hindi, literates and illiterates come to significantly different grammaticality judgments. Since literate and illiterate speakers of a language are equally competent in that language, the etiology of their different judgments can't reside in the competence. It must be learned theory of language driving the difference. Research from developmental psychology provides evidence for the same fact from a different domain. De Villiers and De Villiers (1972; 1974) show that children become grammatically competent well before they make accurate (or at least literate-adult-seeming) acceptability judgments, which suggests that the acceptability judgments do not develop in tandem with grammatical competence itself.

In fact, research on childhood literacy mirrors everything we just canvas on adult literacy and metalinguistic judgment. The same effect of the theory of language acquired while learning to read is observable in studies comparing preliterate with newly-literate children. These studies cover the same sort of task, so I won't discuss them in detail, but examples of the ongoing research program include Hakes (1980), Ryan and Ledger (1984), Adams (1990), Sulzby and Teale (1991), Tolchinsky (2004), and Ramachandra and Karanth (2007). The literatures on adult illiterates and on preliterate children are explicitly brought together in Kurvers et al. (2006). Drawing on adult and child subjects competent in a variety of languages including Moroccan Arabic, Rif Berber, Turkish and Somali, Kurvers et al. replicated the effects found in the aforementioned experiments on phonological, lexical, and semantic intuitions. They document significant differences between adult literates and adult illiterates on eight different types of metalinguistic judgments, agreement between adult illiterates and preliterate children on five of the eight, and disagreement between literates and preliterate children on the same five. This result shows two things. First, the well-documented difference between the metalinguistic intuitions of preliterate and newly-literate children can't be accounted for by appealing to the natural development of linguistic competence, since it mirrors the effects of literacy on adults. Second, it shows that the learned theory of language acquired while learning to read has significant effect across multiple domains of linguistic intuition. In short, decades of research on the relationship between literacy and metalinguistic judgment undermines the claim that linguistic intuition is the voice of the competence rather the voice of learned theory of language.

That literacy has a profound effect on metalinguistic judgments is irrefutable, but not necessarily inconsistent with VoC, since it is possible that learning to read actually alters the competence itself. If this were the case, then the differences in linguistic intuitions between literate and illiterate speakers of the language could be explained by differences at the fundamental level at which language is encoded in their minds. Although this explanation would shield VoC from the literacy data, it is empirically implausible. If learning to read really did change the fundamental way the brain processed language, from semantics all the way down to phonology, we would expect to see striking differences in verbal performance, not just in metalinguistic intuitions. We don't see those differences. In fact, one of the main reasons the literacy data is interesting is precisely because the differences in metalinguistic performance do not mirror differences in linguistic performance. The literature on the effects of literacy on metalinguistic judgment thus provides good evidence against VoC and in favor of LTL.

If LTL rather than VoC is true, we would also expect to see differences between the metalinguistic judgments of the folk and judgments of subjects who have learned scientific theories of language. Linguists themselves often argue that their expert intuitions differ significantly from those of the folk. Sometimes they do so to justify appealing to linguists' intuitions alone (e.g. Gleitman and Gleitman 1970). Sometimes they do so to argue against the common practice of appealing only to the intuitions of a handful of linguists rather than employing statistical survey methods on large samples of the population (e.g. Gibson and Fedorenko 2010). That a difference between the intuitions of linguists and the folk actually exists, however, is controversial, and Schütze has observed that "experimental attempts to establish whether such differences actually exist...have been surprisingly few" (1996, 13).

Results of those "surprisingly few" attempts are mixed. The most prominent early experiments on the issue, those of Spencer (1973) and Gordon and Hendrick (1997), documented unequivocal differences in acceptability judgments between linguists and non-linguists. Unfortunately, methodological issues call into question the scope of their findings. In both cases neither the sample sentences nor the population of linguists surveyed were selected impartially, so we should be hesitant in accepting their results unqualified. Similar methodological worries plague more recent experiments, such as Wasow and Arnold (2005), Dabrowska (2010) Gibson and Fedorenko (2010), which also show marked differences between linguists' and non-linguists' intuitions. Although these methodological flaws call into question whether there are *systematic* differences in linguistic intuition between experts and the folk, however, the fact that so many researchers have consistently been able to find examples of difference suggests that expert training must have some effect on metalinguistic judgments.

Some more recent experiments have been conducted more rigorously, but the relevance of their results to the question at hand are more difficult to interpret. Culbertson and Gross (2009) did find that the judgments of the folk differed from those of linguists, but they also found that the relevant dividing factor was *any* training in the cognitive sciences, not knowledge of contemporary syntax. Sprouse and Almeida (2012), on the other hand, found near consensus between the folk and a linguistics textbook on a large number of grammaticality judgments. So although many linguists have been able to find an effect of training in linguistics on particular intuitions in syntax, the two most wide-ranging and methodological rigorous experiments on the question yield mixed results.

These mixed results appear at first glance to undermine the LTL explanation of linguistic judgments, and thus indirectly bolster VoC. Let's take Sprouse and Almeida's (2012) well-designed and informative study first. Drawing a set of sample sentences from *Core Syntax*, a respected introductory syntax textbook[32], Sprouse and Almeida used Amazon Mechanical Turk to gather acceptability judgments on these sentences from several hundred non-linguists. In nearly all cases the grammaticality status assigned by the textbook agreed with the status assigned by the body of participants. Learning scientific linguistic theory, their experiment seems to show, does not influence a speaker's acceptability judgments.

The experiment does not actually aim to demonstrate this conclusion, however, and we should be hesitant to jump to it given the number of other studies (flawed or not) which contradict it. Sprouse and Almeida's experiment is not actually well-suited to speak to the question of whether learning linguistics affects metalinguistic intuitions. Primarily, this is because the experiment does not directly compare the intuitions of individuals with training in linguistics to those of individuals without it. Sprouse and Almeida's analysis examines whether the directionality (grammatical or ungrammatical) of the judgments of the folk subjects *taken as a whole* agree with the judgments in the textbook. The question we are actually interested in is whether *at the individual level* being a linguist significantly affects the chances that a subject will judge a sentence to be grammatical. The directionality of the aggregate group of subjects may agree with the textbook if, for instance, 60% of subjects judge the sentence to be grammatical and 95% of linguists made the same judgment[33]. But a case like this would be evidence for LTL,

[32] Sprouse, Schütze, and Almeida (2013) gathers similar data using sample sentences from the journal *Linguistic Inquiry*. Similar issues apply to this study for present purposes, though we will contrast them in an analysis later in this chapter.

[33] Later in this chapter the precise level of agreement between subjects will become important, and we'll draw on Sprouse and Almeida's useful data to help determine it. Here, though, since we're interested in

not against. Using merely the textbook's grammaticality assignments doesn't provide the sort of data necessary to make this comparison, because we don’t know the level of agreement among linguists, so the experiment doesn’t say much one way or the other about the truth of LTL.

Sprouse and Almeida (2012) includes two experiments, and the second may be less susceptible to this criticism, since subjects often nearly unanimously agreed, and presumably linguists would as well. Near unanimity on these judgments, however, is likely an artifact of the set of sample sentences. For this second experiment they used mostly examples of sentences identified as grammatical in the textbook, and these are mostly simple, uncontroversial examples such as "Who has read the novel?" and "No one expected him to win." In other words, their second experiment tested intuitions about sentences not likely to be at the heart of argumentation in syntactic theory, where more complicated and controversial examples tend to play the important evidentiary role. The difference between Sprouse and Almeida (2012) and studies such as Wasow and Arnold (2005) and Gibson and Fedorenko (2010) appears to be that the latter used examples from controversial arguments in the syntax literature. Sprouse and Almeida's experiment, while relevant to methodological questions in syntax, does not therefore tell against LTL. Consequently, it doesn't tell in support of VoC, even indirectly.

Culbertson and Gross (2009) provide a different challenge to the LTL account of linguistic intuition. They find that lay and expert intuitions do differ, which seems to support LTL, but there's a hitch. LTL would seem to predict that the dividing line between lay and expert be exposure to syntactic theory, but Culbertson and Gross identify exposure to any sort of cognitive science as the major dividing line. Their experiment included four groups—syntacticians, subjects who had taken at least one syntax course, subjects who had taken at least one cognitive science course, and subjects with no cognitive science experience—and the last group show traits significantly different from the first three. From this datum, they draw the conclusion that the differences in

agreement between lay subjects and linguists, and they don’t provide data which can resolve that question, I don’t bother with precise figures.

intuition are the differences "between subjects with and without minimally sufficient task-specific knowledge" (2009, 722). In other words, the factor producing different responses to sample sentences might not be the possession of different learned theories of language, but instead that lay subjects don't understand how psychological experimentation works and so perform the task incorrectly (or at least differently). If this alternative explanation for the data is correct, it undermines the empirical support for LTL and thus weakens the challenge LTL poses to VoC.

A closer look at Culbertson and Gross' data, however, belies the threat to LTL. We can accept that some of the variability in intuitions is explained by subjects misapprehending the task. In particular, it might partially explain why of the four groups tested, only the group with no exposure to cognitive science shows large within-group variability. Of course, that within-group variability could also be explained by the possession of different folk theories of language, and it is consistent with LTL either way. The other primary datum driving Culbertson and Gross' interpretation is that the three groups with exposure to cognitive science correlated more closely with each other than with the lay group. LTL is consistent with this datum as well, since the inter-group correlations, while strong, are not perfect, and intra-group correlations are also strong. If LTL is true, we would expect to see this. Since members of each group possess similar learned theories of language, we would expect strong intra-group correlations, and since the theories of language taught in syntax and introductory cognitive science courses are among those held by syntacticians, we would expect strong but not near-perfect inter-group correlations. We would also expect that the cognitive-science-only group would correlate more closely with the lay group than the syntax-specific groups would, and this is the case in Culbertson and Gross' data. So on close examination Culbertson and Gross (2012) supports the claim that the theories of language a speaker knows affect her linguistic intuitions.

Let's take stock. We've juxtaposed two hypotheses about the etiology of metalinguistic intuitions. The Voice of the Competence (VoC) account claims that intuitions are produced by the same mental processes which produce linguistic performance. The Learned Theory of Language

(LTL) account, on the other hand, holds that intuitions are unconsciously shaped by knowledge of theories of language, not knowledge of language itself. We have noted the absence of direct empirical evidence for VoC, and adduced two types of empirical evidence in favor of LTL: learning to read and learning linguistic theory both change a speaker's metalinguistic judgments. Therefore, given our current state of knowledge, LTL is the more plausible explanation of where linguistic intuitions come from.

It would be too strong to say that this analysis shows that VoC is false. It is consistent with the evidence cited in this section that linguistic competence and learned theory of language are both factors in the formation of linguistic intuitions. Even if this is the case, however, if LTL is a major component of the etiology of intuitions, it renders problematic any VoC-based defense of the scientific status of linguistic intuitions. According to the literacy and expertise data, LTL provides more than just noise—it is a principle cause of metalinguistic judgments. Consequently, even if the faculty of language contributes in some way to linguistic intuitions, we would need to search for its contributions among everything contributed by LTL and other irrelevant factors.

Some authors argue that we can do just that. Scientific data is almost always messy, and sifting through irrelevant factors is part of the scientist’s job description. We shouldn’t be surprised if gathering evidence through linguistic intuition requires the same sort of fighting through the noise to find the relevant data. Maynes, for example, argues that linguists can “disentangle the complex causal chains” behind a particular intuition “by calibrating intuition through comparison with other data sources” (2012: 459). In a slightly different vein, Hornstein (2015) defends the use of acceptability judgments on the grounds that “(t)hough we don’t have a *general* theory of acceptability judgments, we have a pretty good idea what factors are involved and when we are careful we can control for these and allow the grammatical factor to shine thorough a particular judgment.”

Let’s address calibration first. Maynes considers calibrating of two different sorts. The first is calibration of intuitions with other intuitions, but this will be insufficient to respond to the etiological concerns raised by LTL, because intuitions could converge because of shared learned theory rather than because of shared competence. More promising is calibration of intuitions using usage and processing data to determine when intuitions are reliable. In this case, however, it's not clear what the use of intuitions is adding. If we have to check them against more reliable data anyway, then we may as well rely on that alternative data for our evidence the whole way through.

How about control? It's true that controlling for irrelevant factors is an important part of evidence-gathering in general, but scientific control is only possible under certain conditions. One such condition is if we have a statistical model of how irrelevant factors skew the data, in which case we mathematically correct for those factors post hoc. But we have no such model, because of our limited understanding of how factors like LTL, performance error, social factors, and so on influence linguistic intuitions. Our understanding *that* they have a significant effect does not translate into a statistically sophisticated understanding of *what* that effect is. Another means of scientific control is manipulation—experimental control in the classic sense. If we can intervene during data-gathering to prevent irrelevant factors from influencing the outcome, we can isolate the relevant features (in this case the contributions of the language faculty). For this to succeed, however, we need to both understand the sources of interference, and have the ability to manipulate them. Neither condition is met in the case of linguistic intuitions. Again, our understanding of exactly how different factors contribute to the production of linguistic judgments is limited, which prevents us from identifying where intervention needs to occur. Worse, we lack the ability to manipulate many of these causal factors to the extent needed. We can't just sit a subject down and ask them to give us their acceptability judgment, but ignore LTL, sociolinguistic factors, and semantic features. True, we might be able to train subjects to better avoid some pitfalls, such as rejecting a sentence on the basic of semantic anomaly rather than syntactic ungrammaticality, but we can't do this across the board. Moreover, training of this sort risks producing convergent intuitions by introducing new irrelevant factors, in that the training itself might strongly skew judgments. Neither calibration nor control is likely to allow us isolate the voice of the competence in linguistic intuitions, at least not given present understanding of the psychological mechanisms underlying them. The evidential worth of linguistic intuitions thus can’t be vindicated on the basis of their causal connection to the language faculty.

Reliability

Grant me for the sake of argument that fruitfulness is not a sufficient guarantor of epistemic quality. Grant me as well that what we know about the etiology of linguistic intuitions fails to assure us of their scientific relevance. Neither concession matters if it can be shown that intuitions tend to be right. If intuition-data, that is, is informative and highly reliable, then it is good evidence despite worries about its origins.

To show that linguistic intuitions are reliable, we need to show that they tend to agree with sources of data known to be reliable on etiological grounds, such as usage data and processing data. Obviously, experimental and observational methods of gathering usage and processing data can be done poorly, but when done correctly these methods are non-controversial sources of reliable evidence[34]. We can determine whether intuitions mesh with accepted evidence by exploring whether metalinguistic behavior (intuition) meshes with the linguistic behavior analyzed through experiment and observation[35].

The mesh, I claim, is relatively poor. Metalinguistic judgments frequently entail predictions contradicted by experiments or usage data. This is necessarily so, given the wide variability in linguistic intuitions among competent speakers of the same language. If LTL is true, and we've seen good reason to think it is, we wouldn't expect intuitions to coincide with usage. Learning to

[34] Their reliability at least, is non-controversial. Some linguists question their scope, a concern I will address later.

[35] The previously cited work by Sprouse and Almeida (2012) and Sprouse, Schütze, and Almeida (2013) is sometimes claimed to show that intuitions are highly reliable on these grounds. Their method, however, is to compare intuitions of the folk to intuitions of (single) linguists, and showing a mesh between some intuitions and other intuitions doesn’t demonstrate that intuitions are accurate. We could all have the same intuition and still all be wrong. This is why to validate intuitions we need to check them against non-intuition evidence.

read doesn't substantially change speakers' phonology, but it does change their meta-phonological judgments. Learning some syntactic theory doesn't substantially change the structure of speakers' utterances, but it does change their grammaticality judgments. It follows that usage and metalinguistic data from a population will not coincide. Which theories of language a speaker has internalized is not the only cause of between-subject metalinguistic variation, of course. Schütze (1996) Ch. 4 reviews many contributing factors, but the short version is that metalinguistic variation outstrips variation in linguistic performance nearly every time the issue is studied. He focuses on syntax, and I’ll present data showing similar results in semantics later in the chapter.

No one has attempted the probably impossible task of pinning down exactly how frequently metalinguistic judgments disagree with patterns of usage, but linguists who go looking for examples have no trouble finding them, which suggests that divergence is common. The previously cited Labov (1996) runs through a plethora of examples as do Wasow and Arnold (2005). Looking at a couple will illustrate the general pattern of how intuitions fail.

The first is cited in Wasow and Arnold (2005). Jackendoff (1997) uses as evidence for a claim about idioms the intuition that 'raise hell' is syntactically inflexible, a reasonable intuition. Riehemann (2001) put the claim to the test, using past issues of the *New York Times* as a corpus. He found that 'raise hell' is not only syntactically flexible, but perhaps even fully productive. Examples of usage data included "how much hell can you raise," "All that was really raised was a little hell," and "hell was raised," among others. Any claim based on Jackendoff's original intuition would have been unreliable.

A second example comes from the literature on code-switching. Code-switching is a phenomenon involving a speaker changing languages in the course of discourse, sometimes even within a single sentence. Syntacticians, working largely intuitively, had posited that speakers could not code-switch at certain syntactic junctions, such as between a determiner and its noun, a position still endorsed in, for instance, Belazi et al. (1994). In other words, we might expect an English-Spanish bilingual to say 7 or 8 but not 9 or 10

7. We're going to la playa.

8. Vamos a the beach.

9. We're going to the playa.

10. Vamos al beach.

Whatever we intuitively expect, as it turns out constructions like (9) and (10) are quite common, perhaps even more common than (7) or (8). Sankoff and Poplack (1981, 7) cite some early usage data showing that switches at determiners are possible, and a recent analysis of code-switching on Twitter (Lignos and Marcus 2013) not only decisively showed that determiner-switches are common, but also found patterns in determiner use which would have been difficult to recognize from metalinguistic reflection alone.

A handful of examples, even if they are broadly representative, doesn't prove that linguistic intuitions are always unreliable. They do exemplify a couple of important points, however. First, intuitions can frequently be unreliable, and we can't predict beforehand which intuitions are good evidence and which are not. Second, if intuitions and usage data conflict, we accept the usage data and discard the intuition. So metalinguistic judgments are at least frequently unreliable—much more unreliable than usage data—but this is consistent with their being evidence for language, albeit weak evidence.

But as I demonstrated in the previous chapter, weak evidence frequently fails to meet the general standards of scientific evidence. I'll briefly recap the claim. Consider, for example, anecdotes from sources we believe to be honest. Suppose a pharmaceutical researcher is informed by a trusted relative that this relative found a certain folk remedy to be effective. Testimony from this relative is probably evidence in the strict sense, since it is slightly more likely that the remedy is effective given that testimony than otherwise. Nevertheless, the researcher wouldn't even consider adding her relative's testimony to a scientific article, grant submission, or application for FDA approval. Anecdotal evidence is not accepted as scientific evidence. Similar, appeals to common sense or popular belief, while sometimes evidence in the strict sense, are not acceptable scientific evidence. Nearly half of Iceland inclines towards belief in elves (Sontag 2007), and it is probably the case that elves are more likely to exist given this popular belief than otherwise, but serious science does not accept such popular beliefs as evidence. It therefore remains a possibility that even if linguistic intuitions are a weak sort of evidence they might belong to the class of evidence which is proscribed in scientific practice.

The general norm for evidence in science, I argued, is that to be acceptable scientific evidence, a type of evidence must either

(a) be highly reliable in an absolute sense,

or

(b) be among the most reliable sources of evidence available.

For a full argument in favor of this norm, I refer the reader to the previous chapter, but I give a brief summary here. First, my proposal generally seems to fit the standards of evidence accepted in scientific practice. Introspection, for example, is not generally considered to be a particularly reliable sort of evidence in psychology, but a few sub-disciplines still make heavy use of it. In these sub-disciplines, such as studies of the contents of conscious experience, introspection seems to be ineliminable (Hatfield 2005). This seems to be because, despite its unreliability in an absolute sense, introspection remains the best sort of evidence available for the subject matter of these sub-fields. In other words, because introspection meets condition (b), it is acceptable for scientific use in a few domains despite its fallibility. Scientists, I'm suggesting, tend to appeal to the best sorts of evidence, and sometimes those best sorts are highly fallible.

Additional support for the criteria I've proposed comes from the fact that there are good reasons to exclude from scientific practice evidence which fails to meet both (a) and (b). For an ideally epistemically rational agent, it might make sense to take into account all available evidence, even weak evidence, but scientists are limited beings. For one thing, researchers don't have unlimited resources and can't chase down every last item which might constitute evidence in the strict sense. Proscribing the use of evidence which is significantly outclassed by another source of evidence can thus lead to more efficient allotment of scientific resources. Similarly, good science is hard work. If less reliable evidence is both accepted by the scientific community and easier to obtain than more reliable evidence, scientists will have disincentive to pursue the better sort of evidence. The community as a whole will consequently produce work of less epistemic value. Finally, scientists are probably as likely as the rest of us to overrate certain sorts of evidence, leading to poor scientific inferences. In the case of intuitions this effect is particularly pernicious, since we are especially likely to unquestioningly accept our own intuitions much more than is warranted. Disallowing the appeal to relatively unreliable evidence significantly mitigates this common error in reasoning. In short, from the standpoint of good scientific practice a number of reasons speak in favor of taking something like (a) and (b) to be the criteria for acceptable evidence.

With that groundwork in place we can ask whether linguistic intuitions meet (a) or (b). We've seen that metalinguistic judgements are frequently and unpredictably unreliable, so they fail to meet (a). But they also fail to meet (b). In demonstrating that intuitions can be unreliable, we compared intuition data to data from corpus analysis, psycholinguistic experiment, and sociolinguistic field research. If the intuition data diverged from the data from one of the latter sources, we took this to undermine the intuition data. In other words, we already accept that in terms of evidential reliability intuitions are significantly outclassed by usage and processing data where they overlap in scope. Linguistic intuitions thus meet neither (a) nor (b) except in the probably rare case where we have no other relevant evidence, so regardless of whether they are evidence in the strict sense, they are not the sort of thing which is generally accepted as scientific evidence.

A key premise in this argument is that the evidence from metalinguistic judgments is significantly outclassed by evidence from first-order language use as gathered in corpora, field research, or psychological experiment. Traditionalists about linguistic methodology have occasionally argued against this premise.

Perhaps the most common argument along these lines claims that usage data is problematic because of performance error. Ludlow, for instance, observes that corpora "are not free of error, and can even be deceptive about the linguistic phenomena that we are interested in" (2011a, 68). Fitzgerald insists that corpus analysis is reliant on intuitions, because "[t]he mere occurrence of an expression by itself doesn’t tell you about its grammatical properties. If one wants to know how it is structured, that requires speakers making judgements about its acceptability and interpretation" (2010, 154). The idea seems to be that factors having nothing to do with linguistic competence often lead speakers to produce utterances violating the rules of their grammar. An utterance tokened in a corpus might therefore be an ungrammatical result of a performance error, so we can't use presence in a corpus as evidence that an utterance is grammatical.

It's certainly true that some tokens in any given corpus will involve performance errors, but it doesn't follow that corpora are inferior to intuitions as a guide to grammaticality. Primarily this is because metalinguistic judgments are at least as subject to performance error as linguistic performance is. "[M]entalists do not maintain that linguistic intuitions are the product of linguistic competence alone," acknowledge Maynes and Gross, "other ‘cognitive systems’ not specific to the language faculty also play a causal role. Examples include various sources of ‘noise’ or ‘performance error’" (2013, 719). So the mere fact that usage involves performance error does not affect the relative reliability of usage data vis-à-vis intuitions, since intuitions are subject to performance error as well. Moreover, it is false that the existence of performance error nullifies our ability to determine grammaticality by means of presence in a corpus. We can be confident that a construction tokened relatively frequently is grammatical. Suppose a performance error occurred frequently and systematically, such that it were widely tokened in a corpus. In this case the "error" would be no such thing; we have no grounds to say that a frequent and systematic constituent of language use is ungrammatical merely because it violates our preconceptions about grammaticality. To do so would be to force the data to fit the theory rather than the other way around. Performance error thus gives no ground for calling into question the utility of usage data for syntactic theory.

A similar sort of objection calls attention to types of data corpora seem to be unable to provide. Ludlow claims that intuitions, unlike corpora, give us access to "negative information" such as unacceptable sentences or sentences which would rarely if ever be uttered in actual communication (2011a, 68). This objection underestimates our ability to gather implicit negative evidence from corpora. In a large enough corpus absence of evidence really is evidence of absence. Recall our earlier example of code-switching between the determiner and noun. If Lignos and Marcus (2013) had found only a very few examples of determiner switching among the thousands of code-switched sentences on Twitter it would have been strong evidence that determiner switching is ungrammatical. Corpora, it turns out, give us access to negative information rather straightforwardly. Of course, even if a corpus analysis can tell us *that* a construction is ungrammatical, it can't straightforwardly tell us *why*. On the other hand, we do have some intuitions about why a particular sentence is unacceptable. Wasow and Arnold (2005) call these *secondary intuitions*, and introduce the terminology in order to argue that secondary intuitions are even less scientifically valuable than primary intuitions. Intuitions thus have no advantage over corpora when it comes to negative information. Additionally, even if corpus analysis were to fall short with regards to some types of negative information, well-designed experimental methods can get at negative questions and are epistemically superior to intuition-pumping.

Intuitions do have one advantage over corpus analysis, field research, and lab experiments, however. Intuitions are about as easy and inexpensive to gather as is possible.

Granted, cost is an important consideration in scientific practice. If cost were the only difference between intuition data and usage data, the lesser cost of the former would argue decisively in its favor. But usage data is a more valuable sort of evidence, and the costs of gathering it are not prohibitive, so using intuitions as evidence is not justified on the basis of cost. Furthermore, if gathered well, intuition data may be just as costly as usage data. Schütze (1996) gives a number of methodological proposals for how to gather more reliable intuition data. His suggestions include sampling a large number of lay subjects, using iterated pilot trials to carefully design sample sentences and their linguistic context, learning to avoid experimenter-induced biases and order effects, and conducting more interviews rather than written questionnaires. If gathering good intuition data requires taking up all these suggestions, intuitions will be just as costly as experimental data, and probably costlier than statistical analysis of corpora. Intuitions, it appears, are not as cheap as they seem at first glance, and even if they were it would not justify appealing to them instead of usage data.

Another supposed advantage of intuitions relative to usage and processing data is that they include in their scope phenomena outside the bounds of other data sources. In particular, linguistic intuitions tell us about…linguistic intuitions. This matters if we think that part of the task of linguistic theory is to account for linguistic intuitions. With respect to metalinguistic judgments Chomsky writes, "If a theory of language failed to account for these judgments, it would plainly be a failure; we might, in fact, conclude that it is not a theory of language, but rather of something else" (1986, 37). Yalcin (2013) lists five sorts of facts natural language semantics aims to account for. Two of these five are facts about speakers' intuitions; for comparison, only one has to do directly with communication. If these descriptions about the purposes of linguistics are accurate, and part of the task of linguistics is to explain linguistic intuitions, then it does seem to follow that the intuitions themselves must be counted as linguistic evidence.

It strikes me as odd to think that the linguist is accountable for explaining linguistic intuitions. The physicist is not accountable for explaining common folk-physical intuitions, nor is the economist accountable for explaining intuitions about the causes of unemployment. This is not to say that the study of folk science falls outside the bounds of science. Philosophers of science take a scientific interest in folk theories, as do anthropologists. Indeed, I drew on anthropological studies of folk linguistics earlier in this chapter. But just as physicists and economists do no primarily have to worry about folk physics or economics, the linguist should be accountable for explaining the facts of language, not speakers' judgments about language. Perhaps a better analogy is with psychology. We might reasonably think that the study of folk psychology belongs to the field of psychology, and perhaps the study of folk linguistics should likewise belong to the field of linguistics. If this is so, it only carves out a small space in which intuitions can operate as evidence. Intuitions would be acceptable evidence for the study of folk linguistics, but not for the study of syntax, semantics, phonology, etc. So even if we were to grant the implausible claim that linguistics must give an account of metalinguistic intuitions, the reliance on intuitions as evidence for syntax and semantics would be unjustified.

In short, none of the purported advantages of intuitions over usage and processing data has much purchase, so intuitions fail to meet the scientific standard of evidence. It remains open to the intuition aficionado, of course, to assert that linguistics employs different standards of evidence than other sciences. In practice, I have argued, this is indeed the case. But let's return to our starting point. Linguistics is a serious empirical science, and this fact has significant methodological implications for the practice of linguistics. One of these implications is that linguistics is subject to the same standards of evidence as any other science. If actual practice departs from these standards, we must either adjust the practice or give up on seeing linguistics as analogous to other sciences. Which option we go with is in the end up to the practitioners of linguistics, but it seems clear to me that if we are committed to linguistics being an empirical science and not an a priori discipline that excising the use of intuitions as evidence (to the extent this actually occurs) is preferable.

The non-evidential role of linguistic intuitions

I've been hinting that I'm skeptical that linguistic intuitions actually function primarily in an evidential role, and in this section I explain this skepticism and explore its implications.

Consider the following sentences, which are merely the first sample sentences from a few of the authors sampled in Sprouse, Schütze, and Almeida (2013):

11. *Sarah saw pictures of.

12a. *Was kissed John.

12b. John was kissed.

13a. *John tried himself to win.

13b. John tried to win.

My contention is that we aren't really using our intuitions as evidence that 11 is unacceptable or that 13b is acceptable, because we don't need any evidence for either claim. That constructions such as 11 are unacceptable is not in contention. The background theory shared by syntacticians already entails that "Sarah saw pictures of" is not a grammatical sentence of English. The linguist is not trying to prove that "John was kissed" is better English than "Was kissed John." No proof is needed. Instead, they're simply using the sentences to appeal to a part of the shared background theory, which they will build on in further argumentation. Many, perhaps most, supposedly evidential uses of intuitions in linguistics are actually just an appeal to the common ground of background theory, and not actually making use of intuitions to prove that a sentence is acceptable/true-in-a-situation.

The uncontested claims of a background theory need not necessarily be true, only shared. In order for a field to make scientific progress, scientists must delimit a subset of questions as the questions under examination, and assume a set of fixed background commitments. If we didn't do this, we could never test a particular hypothesis against the data, because we could explain away any result by appeal to a connected claim elsewhere in the web of scientific claims (Duhem 1954; Quine 1951). Over time, a question under examination may become settled, and move to the realm of shared background theory, and occasionally the reverse may occur. Early Newtonian physicists may not have treated the question of whether physical space is Euclidian as a question under examination, and assumed the Euclidean nature of space as given. 19th and early 20th century physicists, however, found reason to raise the question again, and found the claim—which had previously been part of background theory—to be wrong.

Most linguistic intuitions, at least in syntax, are actually appeals to the sort of facts that either we all agree on or that are already entailed by existing shared theory. This has several implications both for my broader argument and for the use of intuitions in linguistics.

Most importantly, it means that we can reject the evidential status of linguistic intuitions without also rejecting wholesale existing work in syntax. The grammaticality of sample sentences like "John tried to win." is not generally in question, even when they appear as part of an argument in generative syntax. A practicing syntactician is usually arguing for an extension or revision to existing theory. In doing so, she is licensed to take any of the generally accepted contents of the theory for granted. Appeals to obviously grammatical or ungrammatical sentences can be understood as appeals to accepted theory. If a syntactician uses "John tried to win." as an example of a grammatical sentence, she is licensed to do so because according to accepted theory it is grammatical. Her intuition is no more than a fast, reflexive application of shared background theory to the sentence. So, insofar as work in syntax generally uses intuitions in this benign, non-evidential way, the rejection of intuitions as evidence gives us no reason to call into question existing syntactic theory. Likewise for semantics, though my suspicion is that intuitions are less frequently innocuous in semantics than in syntax.

On similar grounds, if linguistic intuitions about clear cases like "John tried to win." are not really playing an evidential role, then we can avoid making the absurdly onerous recommendation that linguists check every sample sentence they use against a corpus analysis or psycholinguistic experiment. We might worry that any calls for revision to scientific practice must be moderate, lest the costs of revision be too burdensome. Calling for a group of linguists to abandon their practice of using sample sentences backed by intuitions alone might be too burdensome by these lights. But insofar as the grammaticality of those sentences isn't really in question, my account doesn't call for such a radical revision to practice. Additionally, even if we proscribe the use of intuitions as evidence, intuitions could still play fruitful roles in hypothesis generation and speculative theory-building, and much of what generative syntacticians and formal semanticists do arguably falls under one or the other of these activities. And the fact that many linguists, even those sympathetic to generative syntax and formal semantics, already employ various sorts of usage and processing data as their primary sources of evidence further mitigates the methodological impact of my position. Taken together these considerations show that accepting my conclusions would not entail undue burdens on the discipline.

Finally, this account of the non-evidential role of intuitions has an implication for when the use of intuitions is sound practice. Intuitions, I have suggested, are often merely appeals to the common ground of background theory. Although to be part of background theory in this sense a claim need not be true, it does need to be shared. If the scientists in a field don't agree that a claim is true, or at least a good working assumption, then treating it as in the common ground of background theory will lead to misunderstandings and researchers talking past each other. In cases where there is substantive disagreement in the field, we should therefore try to avoid treating a claim as a part of background theory, and assume falsely that requires no argument or evidential support. Appeals to intuition, then, are only justified when there is broad agreement on the acceptability or truth-conditions of the sentence(s) in question.

The degree to which I'm actually calling for revision to scientific practice, then, depends on the degree to which the intuitions linguists appeal to are broadly shared. To determine this, we need both a measure of sharedness, and data to measure. For a measure, we can adapt a classic measure of diversity in linguistics, Greenberg's (1956) Diversity Index (LDI). The LDI is simple: the linguistic diversity of a population is equal to the probability that two randomly selected individuals from the population will speak different languages. A monolingual population rates 0 on the LDI, and a population where no two individuals share a language rates 1. Papua New Guinea, with its hundreds of languages, has an LDI of .990. Poland, where nearly everyone speaks Polish, has an LDI of .060. Of course, in determining whether linguistic intuitions are broadly enough shared to be reliably considered part of background theory, we want need to invert the measure. We want to determine not the diversity, but the convergence of intuitions. Define *judgment convergence*, then, as the probability that two randomly selected members of a population will share the dominant judgment.

How much judgment convergence should we expect before we're willing to treat an intuition as potentially part of shared background theory? There probably isn't a precise answer, and it probably varies between different scientific contexts, but something in the range of .85-.90 strikes me as a conservative estimate, and I'll use those numbers in assessing the data.

For syntax, the data will come from the aforementioned studies by Sprouse, Schütze and Almeida. Sprouse and Almeida (2012) compared the absolute and relative acceptability judgments of 240 lay subjects to the judgments reported in a popular textbook, *Core Syntax*: *A Minimalist Approach* by David Adger. Sprouse, Schütze and Almeida (2013) involves 936 subjects completing similar tasks on sample sentences they drew systematically from the journal *Linguistic Inquiry*. In appendices to both papers, they include data on individual items—data we can use to assess the judgment convergence of the folk[36] on those sentences/sentence pairs.

Using only their data on binary tasks, judgment convergence (JC) on those

[36] While it would be interesting to assess judgment convergence among linguists, I know of no study as extensive as Sprouse and Almeida's looking at linguists themselves. For present purposes, we will have to assume that judgment convergence among linguists is similar to that among the folk.

sentences/sentence pairs (items) is reported in the table below:

	JC **across items**	**Items with** **JC > .85**	**Items with JC >.90**
Journal Data (152 items)	82.9%	67.1%	55.9%
Textbook Data (250 items)	90.2%	78%	72.4%

Two features of this data are worth calling attention to. First, judgment convergence in syntax is quite high. Most people agree about most sentences most of the time, though there are a significant number of sentences with substantive disagreement. This suggests that, even if I'm right that intuitions are not good scientific evidence for linguistics, we don't need to make major revisions to practice in syntax.

Second, judgment convergence seems to be higher for sample sentences in textbooks than those in recent journal articles. This is precisely what we would expect, if what judgments are generally doing in accessing shared background theory. A textbook will mostly contain the established commonplaces of the field—the shared theory—but journal articles are meant to push the boundaries of theory, so the cases they present are more likely to be controversial. Additionally—and this is merely my subjective impression, so take it with a grain of salt—the textbook sample sentences tend to be straightforward sentences in the vein of "John tried to win," whereas journal sentences are more likely to be syntactically complex and difficult to judge, such as "The dog that I saw's collar was leather." And "Angela wondered how John managed to cook, but it’s not clear what food.” These are actual examples of sentences used in Sprouse, Schütze and Almeida (2013), and it makes sense that researchers trying to push theoretical bounds would need to make use of more complex examples, because the questions that can be answered with straightforward cases are often already settled. The lower degree of judgment convergence on these sorts of sentences, however, should instill in us some caution in expecting that our intuitions are accessing some sort of common ground in these cases. Only slightly more than half of the samples from journal articles hit a judgment convergence level greater than .90, so there is a real need to use more actual evidence when drawing on complicated examples to build on the frontiers of syntactic theory.

So, if this data is representative, many intuitions in syntax are successfully playing a non-evidential but still useful role, but there is a real benefit to be gained from using more usage and processing data for questions on the theoretical edge. Does the same hold true in semantics? Unfortunately, no one has done a study with Sprouse and Almeida’s systematicity and scope on intuitions in semantics. Accordingly, I had to gather data on semantic intuitions myself. While practical constraints limited my scope, I aimed to be systematic, and modeled my methodology on Sprouse and Almeida, drawing items systematically from articles in the journal *Semantics and Pragmatics*.

I’ll present the results on judgment convergence in semantics here, and leave the details of the experiment to the Appendix:

	JC across items	**Items with JC > .85**	**Items with JC >.90**
Semantics: (12 items)	58.0% (95% c.i.: 48.5% –68.4%)	16.7%	16.7%

The situation in semantics doesn't parallel that in syntax[37]. The vast majority of the items don't reach high levels of convergence on truth-conditional judgments. It is not the case, however, that subjects just never agree on the truth-conditions of a sentence. In addition to the items from *Semantics and Pragmatics* I included items I concocted to see how much convergence we get on simple, straightforward cases (these items were not included in the analysis presented in the table above). For example, one such item read:

> "Jill doesn't believe in unicorns and has never seen one. She tells a friend 'I saw a real live unicorn today.'"

All subjects agreed on the falsity of Jill's statement in that scenario. The sample sentences drawn from the journal articles, however, were never so straightforward, and received correspondingly diverse responses from participants. If this result holds true more generally, then intuitions in semantics are almost never innocuous appeals to a common ground of shared theory. Since they are also not good scientific evidence, this gives us good reason to both to treat work relying on semantic intuitions skeptically, and to pursue alternative sources of evidence to make semantic claims.

Conclusion

We began with a trilemma:

(1) If a discipline relies on intuition as a primary source of evidence, it is an a priori discipline, and not a science.

(2) Some subfields of linguistics—notably formal syntax and formal semantics—rely on intuition as a primary source of evidence.

(3) Linguistics is a science and not an a priori discipline.

[37] At least insofar as this data is representative. I would like to expand the number of items and subjects in future work.

At first glance, linguistic intuitions seem to provide some hope for rejecting (1). If our linguistic intuitions are merely reflexively accessed reports of gathered data, or if they are internal reports of the workings of the language faculty, then they are empirical data about language. Unfortunately, in neither of these cases do we have good evidence that intuitions have an empirical relationship to language to the degree necessary to function as scientific evidence.

Given this, to escape the trilemma, we must reject (2). Intuitions frequently play a non-evidential role in linguistics, functioning as shorthand for an appeal to shared assumptions or established theory. An assessment of the extent to which intuitions converge shows that this is a plausible interpretation of what goes on in much of syntax, though work at the theoretical limits of syntax is likely to go beyond where shared assumptions and accepted theory can take us, and thus requires actual empirical evidence of some sort. My outsider's impression of the field, confirmed in discussions with generative syntacticians, is that there is a slowly increasing use of empirical data, particularly neurolinguistics data. If my arguments are sound, this should be celebrated and encouraged. Even if intuitions in syntax often function in an innocuous non-evidential role, they must sometimes be functioning as evidence, either explicitly or implicitly, and this is not sound practice. Using processing data and usage data to evince theoretical innovations is therefore a good idea, even if many or most of the uses of intuitions in syntax are consistent with linguistics' status as an empirical science.

The situation may be different in semantics, which is perhaps unsurprising given formal semantics' continuing close ties to philosophy of language. While we wouldn't want to draw any strong conclusions from the limited data I was able to gather, the frequency of disagreement over the truth conditions of the type of sentences used in formal semantic argumentation suggests that truth-conditional judgments are not only not good scientific evidence, but also not safe appeals to established theory or shared background knowledge. If this is the case, it is even more important for semanticists than for syntacticians to make use of empirical evidence to support their theoretical claims. In fact, if things are as bad as the limit data I have gathered makes them look, formal semantics may indeed often be more of an a priori discipline than a science, and one which we should expect to make slow progress given the conflicting intuitions of its practitioners. This outcome isn’t inevitable. More work on semantic intuitions could help us learn how to distinguish cases where they can play a safe non-evidential role from cases where they function as bad evidence. This knowledge, combined with greater use of usage and processing data in cases where we need actual evidence, could secure semantics’ position as an empirical science. Whether or not that would sever its close ties to philosophy is an open question.

References

Abrusán, M., and Szendrõi, K. 2013. "Experimenting with the King of France: Topics, verifiability and definite descriptions." *Semantics and Pragmatics* 6: 1-43.

Adams, M. 1990. *Beginning to read. Thinking and learning about print*. MIT Press

Adrian, J., Alegria, J., & Morais, J. 1995. "Metaphonological abilities of Spanish illiterate adults" *International Journal of Psychology* 30 (3): 329-353.

Atanassov, D., Schwarz, F., & Trueswell, J. C. 2013. On the Processing of" might". *University of Pennsylvania Working Papers in Linguistics*, 19(1), 2.

Belazi, H., Rubin, E., and Toribio, A. 1994. "Code-switching and X-bar Theory: the Functional Head Constraint." *Linguistic Inquiry* 25 (2): 221-237.

Bertelson, P., Gelder, B. D., Tfouni, L. V., & Morais, J. 1989. "Metaphonological abilities of adult illiterates: New evidence of heterogeneity." *European Journal of Cognitive Psychology* 1 (3): 239-250.

Bloomfield, Leonard. 1933. *Language*. Holt.

Bock, K., Loebell, H., & Morey, R. 1992. From conceptual roles to structural relations: bridging the syntactic cleft. Psychological review, 99(1), 150.

Carroll, Lewis. 1894. "A logical paradox." *Mind* 3 (11): 436-438.

Chomsky, Noam. 1975. *The Logical Structure of Linguistic Theory*. Plenum.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use.* Praeger.

Collins, John. 2008. *Chomsky: a Guide for the Perplexed*. Continuum.

Culbertson, Jennifer, and Gross, Steven. 2009. "Are Linguists Better Subjects?" *British Journal*

Phil. Science 60: 721-736.

Dabrowska, Ewa. 2010. "Naive v. expert intuitions: an empirical study of acceptability judgments." *Linguistic Review* 27: 1-23.

Dahl, Östen. 1979. "Is linguistics empirical?" In Perry, T. (ed.) *Evidence and Argumentation in Linguistics*. 133-45.

Davies, Mark. 2014. "The importance of robust corpora in providing more realistic descriptions of variation in English grammar." *Linguistics Vanguard.*

De Villiers, P. A., & de Villiers, J. G. (1972). Early judgments of semantic and syntactic acceptability by children. Journal of Psycholinguistic Research, 1(4), 299-310.

De Villiers, J. G., & De Villiers, P. A. (1974). Competence and performance in child language: Are children really competent to judge?. Journal of Child Language, 1(01), 11-22.

Devitt, Michael. 2006a. *Ignorance of Language*. Oxford UP.

Devitt, Michael. 2006b. "Intuitions in Linguistics." *British Journal Phil. Science* 57: 481-513.

Devitt, Michael. 2010. "Linguistic Intuitions Revisited" *British Journal Phil. Science* 61: 833-865.

Devitt, Michael, and Sterelny, Kim. 1987. *Language and Reality: an Introduction to the Philosophy of Language.* MIT.

Duhem, P. 1954. (1906) *The aim and structure of physical theory*. Trans. Philip P. Wiener. Princeton: Princeton University Press.

Fitzgerald, Gareth. 2010. "Linguistic Intuitions." *British Journal Phil. Science* 61: 123-160.

Gibson, Edward, and Fedorenko, Evelina. 2010. “The Need for Quantitative Methods in Syntax and Semantics Research.” *Language and Cognitive Processes* (2010): 1–37.

Gleitman, Henry, & Gleitman, Lila. 1979. "Language use and language judgment." In C. Fillmore, D. Kemler & W. Wang (eds.), *Individual differences in language ability and language behavior*. Academic Press.

Gombert, J. 1994. "How do illiterate adults react to metalinguistic training?" *Annals of Dyslexia* 44: 250-269.

Gordon, Peter, and Hendrick, Randall. 1997. "Intuitive knowledge of linguistic co-reference. *Cognition* 62: 325-370.

Greenberg, J. H. 1956. The measurement of linguistic diversity. *Language*, 32(1), 109-115.

Haegeman, Liliane. 1994. *Introduction to Government and Binding Theory*, 2nd ed. Blackwell.

Hakes, David T. 1980. *The Development of Metalinguistic Abilities in Children*. Springer- Verlag.

Harris, Zelig. 1954. "Distributional Structure." *Word* 10: 775–93.

Hatfield, Gary. 2005. "Introspective evidence in psychology." In P. Achinstein (ed.) *Scientific Evidence: Philosophical Theories and Applications.* Johns Hopkins Press.

Jackendoff, Ray. 1997. *The Architecture of the Language Faculty*, MIT Press.

Hornstein, N. (2015, September 15). Faculty of Language. Retrieved February 20, 2016, from http://facultyoflanguage.blogspot.com/2015/09/judgments-and-grammars.html

Karanth, P., & Suchitra, M. G. 1993. "Literacy acquisition and grammaticality judgments." In *Literacy and language analysis* 143-156.

Kolinsky, R., Cary, L., & Morais, J. 1987. "Awareness of words as phonological entities: The role of literacy." *Applied Psycholinguistics* 8: 223-232.

Krifka, Manfred. 2011. “Varieties of semantic evidence.” In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics. An international handbook of natural language* *meaning*, 242-267. Berlin: Mouton de Gruyter.

Kuhn, T. S. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. University of Chicago Press.

Kurvers, J., Hout, R. V., & Vallen, T. (2006). "Discovering Features of Language: Metalinguistic Awareness of Adult Illiterates." In *Low-Educated Second Language and Literacy Acquisition: Proceedings of the Inaugural Symposium Tilburg 2005* 69-88.

Labov, William. 1996. "When Intuitions Fail." In L. McNair et al., (eds.). *Papers from the Parasession on Theory and Data in Linguistics*. Chicago Linguistic Society.

Lignos, Constantine, and Marcus, Mitch. 2013. "Toward Web-scale Analysis of Code-switching". Poster at the 87th Annual Meeting of the Linguistic Society of America, January 5, 2013.

Ludlow, Peter. 2011a. *The Philosophy of Generative Linguistics*. Oxford UP.

Ludlow, Peter. 2011b. “Semantic Knowledge.” In S. Bernecker et al., (eds.). *Routledge Companion to Epistemolog*y. Routledge.

Lukatela, K., Carello, C., Shankweiler, D. & Liberman, I. 1995. "Phonological awareness in illiterates. Observations from Serbo-Croatian." *Applied Psycholinguistics* 16 (4): 463-488.

Luria, Aleksandr. 1976. *Cognitive development: Its cultural and social foundations*. Harvard

Maynes, Jeffrey. 2012. "Linguistic intuition and calibration." *Linguistics and Philosophy* 35: 443-460.

Maynes, Jeffrey, and Gross, Steven. 2013. "Linguistic intuitions." *Philosophy Compass* 8 (8): 714-730.

Miller, L., & Ginsberg, R. B. (1995). Folklinguistic theories of language learning. *Second language acquisition in a study abroad context*, 9, 293-316.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. 1979. "Does awareness of speech as a sequence of phones arise spontaneously?" *Cognition* 7 (4): 323-331.

Niedzielski, N. A., & Preston, D. R. (2003). Folk linguistics (Vol. 122). Walter de Gruyter.

Quine, W. V. O. (1951). “Two Dogmas of Empiricism”, Reprinted in *From a Logical Point of View, 2nd Ed.*, Cambridge, MA: Harvard University Press, pp. 20–46.

Riehemann, S.Z. 2001. *A Constructional Approach to Idioms and Word Formation.* Stanford University Dissertation.

Ramachandra, V., & Karanth, P. 2007. "The role of literacy in the conceptualization of words: Data from Kannada-speaking children and non-literate adults." *Reading and writing* 20 (3): 173-199.

Ryan, Ellen Bouchard, and George W. Ledger. 1984. “Learning to Attend to Sentence Structure: Links Between Metalinguistic Development and Reading”. In John Downing and Renate Valtin (eds.) *Language Awareness and Learning to Read.* Springer-Verlag.

Sankoff, David, and Poplack, Shana. 1981. "A formal grammar for code-switching." *Papers in Linguistics* 14 (1): 3-45.

Schütze, Carson. 1996. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistics.* University of Chicago Press.

Sontag, Katrin. 2007. *Parallel worlds: fieldwork with elves, Icelanders and academics*. University of Iceland.

Spencer, Nancy Jane. 1973. "Differences between linguists and nonlinguists in intuitions of grammaticality-acceptability." *Journal of psycholinguistic research* 2 (2): 83-98.

Sprouse, Jon, and Almeida, Diego. 2012. "Assessing the reliability of textbook data in syntax: Adger's Core Syntax." *Journal of Linguistics* 48: 609-652.

Sprouse, Jon, Carson T. Schütze, & Diogo Almeida. 2013. A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001-2010. *Lingua* 134: 219-248.

Sulzby, E. & Teale, W. 1991. "Emergent literacy." In R. Barr et al. (eds.), *Handbook of Reading Research*, vol. 2, 727-758. Longman.

Tolchinsky, L. 2004. "Childhood Conceptions of Literacy." In T. Nunes & P. Bryant (eds.) *Handbook of Childrens' Literacy,* 11-30. Kluwer

Trueswell, J. C., & Kim, A. E. 1998. "How to prune a garden path by nipping it in the bud: Fast priming of verb argument structure." *Journal of memory and language*, 39(1), 102-123.

Wasow, Thomas, and Arnold, Jennifer. 2005. "Intuitions in linguistic argumentation." *Lingua* 115 (11): 1481-1496.

Yalcin, S. 2014. "Semantics and metasemantics in the context of generative grammar." In *Metasemantics: New Essays on the Foundations of Meaning*, A. Burgess and B. Sherman eds. 17-54. Oxford UP

Appendix—Experimental Design

Participants

In January and February 2015, 220 U.S. subjects were recruited and issued a small honorarium through Amazon Mechanical Turk for a Qualtrics-administered survey. Subjects were excluded from analysis for the following reasons: 27 failed to answer correctly two filler questions with obvious answers, included to ensure subjects actually read and considered the task; 59 subjects were not raised in a home where a variety of American English was the primary language spoken. This left data from 134 subjects for analysis.

9% of subjects had no college experience, 68% had at least some college up to a 4-year degree, and 23% had a postgraduate degree. 20% of subjects were between the ages of 18 and 25, 73% of subjects were between the ages of 26 and 54, and 7% were 65 or older. 61% of respondents were male, 38% were female, and 1% selected “other/prefer not to answer” for their gender.

Design and procedure

Survey items involved sample sentences paired with scenarios. Sentences were adopted and sometimes slightly adapted from articles appearing in the journal *Semantics and Pragmatics* from 2010-2014. I selected sentences by the following process: I only drew a sentence from an article if that article relied principally on the linguist’s truth-conditional judgments to make its case. Articles using psycholinguistic experiments, for instance, were excluded, as were articles on issues other than truth-conditional semantics. From each of the remaining articles, I picked the first sample sentence which played a significant role in making the novel argument of the paper (in one case, it was a contrastive pair.) Established example sentences canvassed in a review of extant literature, for instance, would not qualify.

The text of the survey is appended below. The order of items was randomized. For purposes of calculating judgment convergence, the ranges 0-3, 4-6, and 7-10 were treated as classes of equivalent answers.

Instructions

Each question will consist of a short case description ending in a sentence spoken by a character. Rate on a scale of 1-10 how *truthful* each statement is.

. Art watched Kathryn win the race. Art later tells Sarah,

"Floyd won the race, I hear, but I was there and he didn't"?

. Mary, Sally, and Betty are the only female students, but John knows nothing about them. John is looking at pairs of pictures of them (i.e. two pictures of Mary, two pictures of Sally, and two pictures of Betty). Again, he may not be aware that Mary, Sally, or Betty are students, or even female. John mistakenly thinks that each pair of pictures represents two different people. For example, he thinks that the two pictures of Mary are not of the same person, but of two different people. For each pair of pictures, John points first at one member of the pair, then at its second member, and says to himself sincerely, "This person likes that person's mother."

An onlooker observes

"John believes that every female student likes her mother."

.Sharon is explaining to you what she and her friends did last night. She tells you, "We were hungry, but the restaurants were closed." Luke overhears the conversation and later tells his brother that

"Sharon and her friends ended up eating together last night"

This item was not drawn from an article and was not included in the analysis

.Ted believed in Santa Claus as a child, but no longer does. He tells Kelly,

"Santa Claus isn't real."

. A hexagon is a six-sided shape, and Laura knows this. She tells Jill

"A hexagon has at least 5 sides."

.Alissa is explaining to Ann who Alice invited to the party. Ann asks, "Did Alice invite Fred and Sue?" Alissa says

"The people she invited weren't Fred and Sue. She invited Fred, Sue, and Gordon"

.Steve sincerely tells Kris, "If John is a diver and wants to impress his girlfriend, he'll bring his wetsuit." Kris later tell Joanne that

"Steve believes that John owns a wetsuit."

This item was not drawn from an article and was not included in the analysis

. Jill doesn't believe in unicorns and has never seen one. She tells a friend

"I saw a real, live unicorn today."

. Ortcutt is the town mayor. One night, Ralph sees a stranger sneaking around on the waterfront. That stranger happens to be Ortcutt, but Ralph doesn't recognize him; in fact, Ralph has never heard of Ortcutt. Ralph believes that the fellow he sees sneaking around is a spy.

Just for fun, Ralph imagines that the man he sees is flying a kite in an alpine meadow rather than sneaking around the waterfront.

Knowing all this, Jess writes that

"Ralph imagined that he did not see Ortcutt sneaking around on the waterfront."

. Penny sincerely tells Tyler, "Either Robert didn't use to smoke heavily, or he stopped smoking." Tyler later tells people that

"Penny believes that Robert used to smoke."

.Italy is a warm country. Knowing this, Julie tells you

"Some Italians come from a warm country."

.Frank learns that recruits who fail basic training are ejected from the Army. He tells you

"All longstanding members of the Army have passed basic training."

.In a footrace Luke's speed exceeded the speed of at least the slowest student but not necessarily each individual student. Charles announces that

"Luke ran faster than every student did."

[Only 50% of subjects saw the following item]

. Kelly asks Sam, "Have you ever seen horses in Whispering Meadow?"

Sam has only ever seen one horse in Whipering Meadow.

Sam replies

"No"

[The other 50% of subjects saw the following item]

. Kelly asks Sam, "Have you ever seen horses in Whispering Meadow?"

Sam has only ever seen one horse in Whipering Meadow.

Sam replies

"Yes"

CHAPTER 4: SIMULATION IN EVOLUTIONARY LINGUISTICS

Introduction

The evolution of human language has been a perennially controversial topic of research. Notoriously, in the late 19th century the Linguistic Society of Paris banned publication concerning the origins of language, since much of that research consisted of wild speculation. In the early 20th century, Ferdinand de Saussure, the founder of modern linguistics, displayed both a lack of interest in and skepticism towards research on the evolution of language (Robert 2010). Noam Chomsky, during the second half of the 20th century, has been a frequent critic of evolutionary linguistics. This skepticism, persistent at least a century and a half, arises in large part from the paucity of empirical evidence available.

Research into the origin and evolution of language, however, has just as persistently remained a popular subject of research. Recently, evolutionary linguists have begun to supplement the scanty empirical evidence with computer simulations, in particular a type of simulation—agent-based models (ABMs)—which has proven fruitful in other biological and social sciences. It is unclear, however, how this addition to the researcher's toolkit is meant to address the skeptic. Computer simulations run the risk of providing the illusion of understanding while merely giving new form to speculative theories. Is a fruitful, empirically-grounded evolutionary linguistics possible? And do simulations have a role to play in allowing us to make progress on the difficult questions regarding the origins and development of human language?

In what follows I give a tentatively optimistic affirmative answer to both questions. I'll argue that although simulations do not provide evidence on their own, they can extend the inferential reach of weak empirical evidence. This means that by using computer simulation in tandem with data from comparative biology, archeology, historical linguistics, psycholinguistics, and so on, we are actually well-equipped to make progress on a large set of questions about language evolution. On the way to establishing this conclusion I'll draw some general lessons about the role of independent empirical data in agent-based modeling.

This chapter will unfold as follows. I begin by outlining the reasons for skepticism about research into the origins of language, focusing on why the extant empirical evidence is so uninformative. Next, I discuss the arguments for why ABMs and other computational methods are able to remedy the situation. After reviewing commonly-given reasons for doubting that simulation can do so, I turn to defending the use of simulations from these complaints. Addressing these worries will require providing an epistemology of computer simulation, and I review the standard philosophical account, only to argue that it is inadequate when applied to the use of simulations in evolutionary linguistics. Drawing on insights from evolutionary linguists and practitioners of agent-based modeling in nearby fields, I provide a new account of the epistemology of simulation as it is used to study evolutionary processes. On this account, simulations allow us to draw out the inferential implications hidden in the extant sources of empirical evidence. After illustrating my account with concrete examples from evolutionary linguistics, I conclude by discussing the implications of the account for skepticism towards evolutionary linguistics.

Reasons for Skepticism

The case for skepticism towards evolutionary linguistics is straightforward: we don’t have sufficient empirical evidence to craft well-informed, well-confirmed theories. Even authors who work on the evolution of communication systems worry that there are severe constraints on their inquiry. “Emphasis on the past adaptive history of the language faculty is misplaced,” they argue, since “[s]uch questions are unlikely to be resolved empirically due to a lack of relevant data, and invite speculation rather than research” (Fitch, Hauser, and Chomsky 2005). This “lack of relevant data” leads leading figures to declare a near total lack of progress in evolutionary linguistics. The high-profile authors of Hauser et al. (2014), including linguist Noam Chomsky, biologist Richard Lewontin, and paleoanthropologist Ian Tattersall argue that "[b]ased on the current state of evidence...the most fundamental questions about the origins and evolution of our linguistic capacity remain as mysterious as ever." These are strong claims about the futility of the central aims of evolutionary linguistics.

Does the actual state of evidence bear out these strong claims? To some extent it does. If we consider the principle sources of evidence available to evolutionary linguists, we see that each is either weak, scanty, or largely irrelevant to the central questions. Consider the case of paleoanthropology. Language, of course, does not fossilize. So while paleoanthropology in general is difficult due to sparse evidence, paleolinguistics is even more difficult because the evidence we do have reflects rough facts of anatomy and material culture, not the status of human cognition and communication. This is not to say that facts of anatomy and material culture provide no evidence for language evolution. We can, for instance, infer something about the development of modern phonological capacity by studying the bone structure of hominid fossils, but even on these issues there is wide disagreement about what exactly to infer from the fossil record (Jackendoff 1999). Furthermore, evidence of this sort is of limited scope, telling us almost nothing about development of modern systems of syntax or semantics.

Comparative biology, particularly comparison with our near evolutionary relatives, serves as another important source of evidence for evolutionary linguistics. But the data from comparative biology is also sometimes claimed to be of little help to the linguist. “Language appears to be unique to the species H. sapiens,” argue Bolhuis et al. (2014), “That eliminates one of the cornerstones of evolutionary analysis, the comparative method.” Natural human language, the argument runs, is not continuous with the communication systems or higher cognitive capacities of other primates. Research into the behavior of those primates, if this is true, provides little aid in our theorizing about the development of human language.

There are similar reasons to be skeptical about the other principle sources of evidence in evolutionary linguistics. Evidence from historical and comparative linguistics is plentiful, but apparently of limited reach. First of all, the limits of historical reconstruction are only a few thousand years, and that upper limit is only for our best cases. But these best cases (e.g. Indo-European, Afro-Asiatic, and Sino-Tibetan) represent a very small minority of linguistic diversity, and so provide dubious evidence for general principles of linguistic change, let alone information about the evolutionary history of language going back more than a few thousand years. Some have argued that historical and comparative linguistics tells us little about the moderately deep past at all. “I do not consider comparative historical linguistics a branch of prehistory,” Harrison argues, “and I sincerely believe that if we cared less about dates, maps, and trees, and more about language change, there’d be more real progress in the field” (2003: 231). If this skepticism is well-placed, it excludes from evolutionary linguistics its most prolific source of evidence—specific details about the structure of modern spoken and recent written languages.

The last major source of evidence for evolutionary linguists comes from contemporary cognitive science, including experiments with modern human subjects. No one argues, of course, that this sort of evidence is irrelevant to evolutionary linguistics, but we might still worry that facts about the evolutionary endpoint of language evolution tell us little about its origins or evolutionary trajectory.

At first glance, then, the state of the evidence might justify the dogmatic pessimism towards evolutionary linguistics expressed by authors such as Fitch, Hauser, and Chomsky (2005) and Hauser et al. (2014). The evidence is always some combination of sparse, weak, and irrelevant. Of course, the evolutionary linguist can push back on a number of fronts. She might argue, for instance, that primate communication systems actually are continuous with human language in some sense, thus partially vindicating evidence from comparative biology. I’m optimistic about vindicatory strategies of that sort, but I will pursue a different response to the skeptic in what follows. According to the defense I will pursues, we can make good use of even sparse, weak evidence using the tools provided by computational methods.

Simulation to the rescue?

Evolutionary linguists, of course, are aware of the poor state of evidence in their field, but research continues. The pessimism expressed by the skeptics runs counter to the optimism expressed by many scientists who, while acknowledging the scarcity of evidence, think we have the tools to make progress on this difficult subject matter. Among the most important of these tools are computer simulations. In the introduction to a volume on simulations in evolutionary linguistics, Cangelosi and Parisi acknowledge that theories of language evolution “inevitably tend to remain speculative because we don't have the empirical evidence to confirm or disconfirm them or to choose among them,” but argue that “[i]t is this very problematic aspect of the study of language evolution which computer simulations can help us to overcome" (2002: 4-5). In particular, it has been argued that agent-based models allow us to "compensate for the lacking empirical evidence by utilizing methods from computer science and artificial life" (Lekvam et al. 2014: 49).

An agent-based model (ABM), sometimes called an individual-based model, is a simulation which represents "a system's individual components and their behaviors. Instead of describing a system only with variables representing the state of the whole system, we model its individual agents" (Railsback and Grimm 2011: 10). In the case of evolutionary linguistics, this means that an agent-based model is a model in which individual organisms are represented, and make local decisions based on interactions with other organisms and the environment. Alternative forms of simulation would represent, for instance, populations or languages as holistic entities. In an ABM, group-level properties, such as shared languages, must emerge from the features and behavior of individual agents. For this reason, ABMs suit evolutionary linguistics well, since language is a social feature which emerges from facts about individual agents.

Not everyone, however, agrees that computer simulation can address the evidence problem in evolutionary linguistics. It would be surprising indeed if simulation allowed us to pull evidence out of nowhere, as if by magic. Hauser et al. contend on this basis that we can dismiss extant simulations in evolutionary linguistics, noting that "all modeling attempts have made unfounded assumptions and have provided no empirical tests, thus leaving any insights into language's origins unverifiable" (2014: 1). Less polemically, Martins et al. (2014: 84), argue that computer "models should be interpreted as quantitative demonstrations of logical possibilities, rather than as direct sources of evidence," and express concern that evidential uses of simulations are determined by the modeler's favored theory and not by the historical facts. One worry, then, is that simulations are not grounded in the reality of their target systems, so they cannot play a confirmatory role.

A second worry is that the way simulations are used in evolutionary linguistics is fallacious. Discussing the use of simulation to study human evolution, Templeton writes that "The ecological fallacy consists in thinking that the relationships observed for the aggregate data prove that an underlying process model is true… Because of the ecological fallacy, the simulation/goodness-of-fit approach can never be considered a test of any model of human evolution" (2007: 1515). Suppose that we write a simulation in which the aggregate behavior of the agents looks strikingly similar to natural language. Can we infer from this that the mechanisms implicated in the simulation are what actually produced natural language? Templeton's argument is that we could not, since the simulation outputs are multiply realizable. We can never infer from goodness-of-fit of the end result of the model to an observed fact that the underlying processes in the model also fit well with reality.

Gareth Roberts brings up a third worry. In facing the practical problems with studying the cultural evolution of language, he observes that "[s]imulations in *silico* provide a solution to some of the practical difficulties, but, as we have seen, give conflicting results, partly to be explained by the necessary simplicity of the agents that interact in them" (2010: 140). He accurately recognizes that for any particular simulation it would not be hard to produce another which gives a conflicting result. The problem, he suggests, is that the 'agents' in ABMs resemble the complex organisms they are meant to represent very little. In creating abstract agents for computer modeling, we leave out many potentially relevant features of the organisms in question.

Given these three worries—disconnect from reality, fallacious inference, and oversimplification—it's no longer so clear that ABMs really can "compensate for the lacking empirical evidence" in evolutionary linguistics. Nevertheless, I want to push back against these objections. To do so, we'll need an account of precisely how simulations are supposed to help, so we turn next to the epistemology of simulation.

Philosophers on simulation: the prediction account

Although computer simulations have been used in the sciences for decades, most of the philosophical work on their role in science is recent. The most prominent philosophical analysis of simulation, perhaps, belongs to Eric Winsberg. Winsberg (2009) argues persuasively that traditional accounts of confirmation are inadequate to the task of explaining the epistemology of simulation. In part this is because "philosophy of science has traditionally concerned itself with the justification of theories, not their application," and simulations are often the application of theories[38] (2009: 836). It is also due to the fact, Winsberg argues, that simulations involve practices such as idealization and trial-and-error which have not been adequately treated by philosophers of science. Therefore, we cannot just import a general account of confirmation to explain how we learn about the world through computer simulation.

On Winsberg's account, the chief way we learn about the world through simulation is by using simulations to make predictions based on established theory. "The epistemology of simulation," he writes, "is rarely about testing the basic theories that may go into the simulation, and most often about establishing the credibility of hypotheses that are, in part, the result of applications of those theories" (2009: 836-37). In many cases, such as the simulations in physics and climatology which Winsberg takes to be paradigm cases of simulations in the sciences, we have an accepted background theory which doesn't immediately yield concrete predictions. Often this is because the best mathematical model of that theory is a set of analytically unsolvable equations, and this is where simulation comes in. If we can find "a way of implementing that model in a form that can be run on a computer" we can use simulation to make inferences to predictions founded on the accepted theory (2009: 836).

To illustrate, consider how computer simulation is used to predict the weather. We'll use as an example the commonly employed Global Forecast System (GFS)[39]. The GFS is a complex simulation which models weather conditions worldwide. The background theories informing the GFS includes a number of different domains, including fluid dynamics, thermodynamics, soil chemistry and physics, electrodynamics, and so on. Each of these well-established physical theories provides a set of differential equations relevant to the weather, but taken as a whole they

[38] In their discussion of how evolutionary linguistics can make use of computer simulation, Cangelosi and Parisi explain that "a simulation is the implementation of a theory in a computer" (2002: 5), so they might well agree on this point.

[39] Accessible at http://www.emc.ncep.noaa.gov/index.php?branch=GFS

are analytically intractable. So instead of trying to create a monstrous hybrid theory out of all the relevant theories, the GFS instead divides the Earth's surface and atmosphere into a three-dimensional grid, translates each individual theory into an algorithm applicable to that grid system, then uses observational data to set the initial parameters (temperature, ozone content, amount of sea ice, etc.) for each location in the grid. With all this in place, the model predicts weather patterns for 8 days following the initial conditions, and it currently predicts major weather patterns with ~80% accuracy[40].

So why should we expect the predictions yielded by the GFS to be reliable? According to Winsberg's epistemology of simulation, in large part because of the reliability of our theories of fluid dynamics, thermodynamics, soil chemistry and physics, electrodynamics, and so on. The simulation output is reliable because the input is reliable. Of course, that can't be the whole story. Winsberg emphasizes that translating theory into simulation is hardly straightforward, and involves "physical insight, extensive approximations, idealizations, outright fictions, auxiliary information, and the blood, sweat, and tears of much trial and error" (2009, 837). It is important to acknowledge these aspects of simulation for a number of reasons, not least of which is that if a simulation based on established theory yields inaccurate predictions, we shouldn't immediately assume that the problem lies with the theory.

Wendy Parker (2008), builds on Winsberg's epistemology of simulation by addressing how scientists handle the issues introduced by idealization, trial and error, and the other areas of simulation involving the shedding of sweat and tears. She does so by drawing an analogy between simulation and experimentation. The experimenter, Parker observes, does not just perform an experiment, observe the result, and then accept that result as valid given how well her experiment was designed. Instead, she also engages in what Parker calls the "Sherlock Holmes strategy" (following Franklin 1986), which "involves showing that all plausible sources of error and alternative explanations of the result can be ruled out (as unlikely)" (Parker 2008: 174). For an

[40] http://www.emc.ncep.noaa.gov/gmb/STATS/STATS.html

experiment, these sources of error might be a contaminated sample, a misreported observation, or the use of an inapplicable statistical technique. The wise scientist, Parker argues, knows to make sure that none of these sources of error are the actual explanation of a result before putting faith in it. Similarly, the modeler running a computer simulation knows that her simulation's predictions can only be trusted if the plausible sources of error can be ruled out. For a simulation such as the GFS, these sources of error might include bugs in the code, poor implementation of the relevant differential equations as discrete algorithms, hardware limitations, bad parameter values introduced by faulty observers or equipment, failure to account for a significant contributing factor, etc. Our confidence in the predictions of the GFS depends not only on our trust in the physical theories underlying it, but also on our trust that most of these sources of error have been eliminated.

Of course, in a simulation as complicated as the GFS, we wouldn't expect them to be completely eliminated, and the same holds true even for less complex simulations. This leads both Parker and Winsberg to observe that part of the epistemology of simulation involves the calibration and refinement of a simulation over time to improve its accuracy. Sticking with the GFS as our example, this has involved years of updating the simulation[41] in response to bugs found, new elements determined to be significant, and comparison with other similar models. The result has been an increase in accuracy of 20-30% over the past three decades[42]. Even scientists using younger, less ambitious models, however, tinker with them to improve their reliability, and any account of the epistemology of simulation must account for this.

To recap, the emerging philosophical account of the epistemology of simulation emphasizes the epistemic reasons we have to accept the predictions simulations yield. In particular, if a simulation is an instantiation of some aspect of a trustworthy scientific theory, then all else being equal our confidence in the simulation is warranted by our confidence in that theory.

[41]Recent changelogs: http://www.emc.ncep.noaa.gov/GFS/impl.php

[42] http://www.emc.ncep.noaa.gov/gmb/STATS/html/aczhist.html

To verify that all actually *is* equal, we must also rule out plausible sources of error, and scientists do so both by considering those sources individually, and be refining their computer models in response to failures and to new information. Call this epistemology of simulation the *prediction account*, since it explains why we should trust predictions produced by computer simulation.

The prediction account and evolutionary linguistics

Applied to models such as standard weather and climate simulations the prediction account holds up well, and those are the sorts of simulations Parker and Winsberg have in mind. But simulation frequently plays a different role in fields like evolutionary linguistics than it does in the physical sciences. In its emphasis on prediction, the prediction falls short of accurately describing the practicing of simulation in evolutionary linguistics, and thus can't provide us with tools to address the problem of weak evidence.

Agent-based models in evolutionary linguistics generally aim not at prediction but at explanation and theory confirmation. It is this confirmatory aim which, if successful, would allow evolutionary linguists to make progress in the face of scarce evidence, so we'll focus on simulations in evolutionary linguistics which aim to confirm particular theoretical claims about the evolution of language. To get clear on how confirmatory simulations fall outside the scope of the prediction account, let's look at a typical example of inference from simulation in evolutionary linguistics.

Gong's (2011) lexicon-syntax coevolution (LSC) model is designed to address two competing theories on the development of modern language from its protolanguage precursor. One theory emphasizes *synthesis*, hypothesizing that the protolanguage consisted of individual words with semantic scope similar to modern words, and the jump to full language consisted of developing syntactic methods of combining these words (Kirby et al. 2007; Bickerton 2008). The other focuses on *analysis*, suggesting that protolanguage consisted of holophrases—individual words carrying the full meaning of a modern sentence—and that the shift to full language occurred when these holophrases were segmented, thus creating compositional units of meaning (Wray 2002). Both the synthetic and the analytic theories are plausible accounts of the actual emergence of language, but they are inconsistent. Gong's aim, given the "lacking direct evidence" (2002: 67), is to use an agent-based simulation to assess the two theories. Note that this is not a predictive aim, but a confirmatory one.

The LSC model consists of a number of elements. Meaning in the model is represented by elements from a pre-determined semantic space being placed in a predicate-argument structure. For instance, if the semantic space has elements *wolf*, *deer*, and *stalk*, one potential meaning would be "stalk<wolf, deer>" ("wolf stalks deer") and another would be "stalk<deer, wolf>" ("deer stalks wolf"). Meanings are not transmitted directly; signals are strings constructed from a set of initially meaningless syllables. Because the model is agent-based, meanings and signals are manipulated by agents, who have the ability to internally represent mappings from meanings to signals, as well as to represent compositional rules for combining elements of signals. Agents are additionally able to generate new signals where none are available, and, as proposed by the analytic theory, to attribute structure to sub-constituents where pattern has serendipitously emerged. Finally, agents have limited memory, ensuring that agents will abandon semantic and syntactic rules which don't assist in communication. During each round of the simulation, each agent is randomly paired with another member of the population, and the two attempt to communicate. If the communication is successful, each agent reinforces the rules they used to communicate; if communication fails, they penalize those rules, and hypothesize new rules if necessary[43].

Gong uses the LSC model to address a number of theoretical questions in evolutionary linguistics, including the question about whether the emergence of language is analytic or synthetic. In runs of the simulation, at first "exchanged sentences are primarily formed by holistic rules. Given more linguistic experiences, recurrent patterns start to appear and get acquired as compositional rules, and a competition occurs between compositional and holistic rules" with compositional rules eventually winning out (Gong et al. 2014: 285). Gong takes this result to

[43] Precise details can be found in Gong (2011).

support Wray's analytic theory of the origins of modern language. Gong also notes some other interesting results of the model, including that new syntactic rules tend to be applied first to individual words and then to whole categories of words (bottom-up syntactic development) rather than the other way around. Of course, in the end Gong is careful to hedge his claims: they aren't "definitive" but they "may partially explain" language origins (2011: 98).

It should be clear by this point that the prediction account doesn't address the type of inference that Gong uses the LSC model to make. Recall that Winsberg develops the prediction account in response to the fact that scientific simulation is "rarely about testing the basic theories that may go into the simulation, and most often about establishing the credibility of hypotheses that are, in part, the result of applications of those theories" (2009: 836-37). As the LSC model exemplifies, however, the reverse is true is evolutionary linguistics. The "basic theories that may go into the model" are precisely what is in question—in the case of the LSC model, one question is "Can a population of learners who analyze holophrases develop a successful compositional communication system?", so the modeler simulates agents who can analyze holophrases, thus testing the analytic theory of language origins. Contrast the LSC model with the GFS weather model. We have good, reliable theories of physical system, so we can implement them in the GFS model, which inherits their reliability. But we don't have theories of language evolution which are known to be reliable, so the LSC model can't be deriving its inferential strength from the basic theories going into it.

The fundamental problem is that the scientific aims of the models are different. The LSC model doesn't aim at prediction. Nor can we assess the model by comparing its predictions to actual outcomes, as we can with models like the GFS model. Conversely, the GFS model doesn't primarily aim at confirmation or explanation. The whole point of the LSC model, on the other hand, is to confirm particular theories (the analytic theory, bottom-up syntax, etc.), and to provide an explanation for features of natural language. The LSC model thus falls outside the scope of Winsberg's prediction account of the epistemology of simulation. This means that the leading philosophical account of simulation won't provide the tools we need to respond to those scientists who are skeptical that simulation can contribute to evolutionary linguistics.

Simulation as a means to draw out the inferential power of evidence

The problem with Winsberg and Parker's accounts of the epistemology of simulation is that they rely on a simulation's basis in established theory to justify inferences drawn from the simulation, a path generally unavailable to evolutionary linguists. Even though simulations in evolutionary linguistics can't just be implementations of accepted theory, if they are to have any confirmatory power they will need to have some epistemic connection to the facts. Although we don't have anything like an accepted theory of language evolution, comparative psychologist Michael Tomasello notes, that "[t]here are a number of processes currently occurring in the natural world that were very likely involved in the origins and evolution of language. These can be studied empirically, and simulations can be compared to them." In other words, we might not have established theories of language evolution, but we do have some idea about particular features of agents that belong in our agent-based models. Gong justifies his LSC simulation, for instance, in part by arguing that his agents' "learning mechanisms are traced in empirical studies" (2011: 78). If simulations in evolutionary linguistics are able to be even weakly confirmatory, it will be because of their connection to these sorts of facts. The question is how simulation, in tandem with knowledge of some facts of this sort, can confirm or disconfirm hypotheses about language evolution.

The answer, I'll argue, is that the evidential import of known facts can be present but unclear, and simulation can reduce the unclarity. In other words, *simulations can aid in confirming or disconfirming hypotheses about language evolution because they draw out the inferential import of the empirical evidence we already have*. Recall that empirical evidence on the origins of language is typically spotty and weak. We have limited prehistorical data, and it isn't clear what conclusions about language origins we can draw from contemporary data. Simulations can remedy this situation in two ways. First, they can connect the dots between observed facts and particular hypotheses. Second, they can situate contemporary evidence in otherwise hard-to-test evolutionary settings.

As Cangelosi and Parisi observe, theories of language evolution tend to be "stated in vague and general terms," but computer simulations "cannot but be explicit" (2002: 5-6). This contrast underlies the first way in which simulation can draw out the inferential import of weak evidence. Connecting a particular hypothesis to a computer simulation requires a precise operationalization of that hypothesis, either to implement it in the simulation, or to compare it with the detailed output of the simulation. Likewise, implementing an observed feature of natural agents in an agent-based model requires a precise mathematical model of that feature. Such a model leads to both an increased ability to make precise inferences on the basis of that feature and a means of modeling the effect of that feature on other features of interest. This sharpening up of both our hypotheses and our models of the evidence enables us to better connect the dots between evidence and hypotheses. In short, using simulation allows us to better determine how the empirical evidence we have should lead us to adjust our confidence in the various hypotheses.

Simulation can accomplish this not only by making our understanding of hypotheses and evidence more specific, but also by resituating observed evidence in evolutionary settings, which can be difficult to do in the flesh. We can, with some clever planning, conduct laboratory experiments in the evolution of *E. coli* or *D. melanogaster*, but allowing humans to struggle for existence and reproductive success in the lab would be nearly as unwieldy as it would be unethical. Instead, we have over a century's worth of work in theoretical biology which gives us mathematical frameworks for describing evolution, and we can situate data from cognitive science, archaeology, comparative linguistics, etc. in these frameworks through computer simulation. This allows us to draw out the implications of the often weak sources of evidence we have for evolutionary hypotheses.

Some examples will demonstrate how simulations can draw out the inferential implications of the evidence in both these ways. Regier et al. (2015) claim that it is a cross-linguistic fact that the extensions of terms in a number of domains tend to partition the relevant semantic space near optimally. For example, although languages differ widely in the number of basic color terms they employ, field research from the World Color Survey shows that the basic color terms of each language tend to divide up perceptual space (quantified using the Munsell color system) into chunks such that each term is near maximally informative. According to empirical work by Regier et al. similar claims hold true cross-linguistically in domains such as kin naming systems and spatial relation descriptors. Supposing Regier et al. are right, in core semantic domains evidence suggests that given a particular number of terms, a language will partition the semantic space near optimally.

On its own, this fact tells us something about the evolutionary result of language evolution, but that doesn't get us very far in choosing one theory of language evolution over another. Combined with simulation, however, it can. If a simulation of the evolution of a semantic space doesn't produce this sort of optimal semantic partitioning, then we know that, all else being equal, the theory underlying that simulation is flawed. On the other hand, a simulation which successfully outputs such semantic partitioning might weakly confirm a hypothesis underlying a theory of semantic change. To give a toy example, Skyrms (2010) uses simulations to show that basic Lewis-Skyrms signaling games with more than 2 or 3 signals don't reliably evolve optimally informative signals. In tandem with Regier et al.'s empirical data, then, those simulations provide a mark against basic Lewis-Skyrms signaling games as a complete hypothesis of the origins of semantic extensions. By adding speaker cost to signaling games, however, Jäger (2007) successfully simulates the evolution of near-optimal partitions of semantic space. Given the empirical data, this simulation provides a mark in favor of the hypothesis that one factor driving the evolution of semantic spaces is cognitive demands on speakers. These two examples show how simulations can connect the dots between empirical evidence and particular hypotheses, showing which should be confirmed and which disconfirmed by that evidence.

Another example: Xu et al. (2013) performed a lab experiment on the cultural evolution of color terms. They generated random partitions of the Munsell color space, divided into 330 swaths, for different numbers of terms (n=2, n=3, …n=6). The first round of subjects was trained on fabricated color terms using these random partitions as training data. For instance, a subject would be shown a sample of one of the 330 swaths and informed that it was an example of [*fabricated color term*]. After training, the subjects were presented with samples of each swath and asked to name them using the terms they had just learned. Their output was used as the training data for the second round of subjects, who repeated the process. The third round of subjects used the output of the second round as training data, and so on for twelve rounds, thus creating experimental cultural evolution in the lab. The key result of the experiment is that after only five or six rounds, the initially random partitions came to resemble the partitions of color space in natural languages, even natural languages that the subjects did not speak. Xu et al. take this as evidence that cognitive or perceptual biases affect the cultural evolution of language.

Assuming that this is the correct interpretation of the empirical data, what does this result tell us about language evolution? On its own, it tells us merely that biased transmission likely plays a role in semantic change. Simulations, however, can extend the significance of this observation. Knowing that biased transmission is an empirically demonstrated mechanism of semantic change lends credence to simulations of language evolution which incorporate biased transmission, such as those of Kirby and Hurford (2002). Using those simulations, we can explore which sorts of bias and which sorts of transmission produce the effect in more realistic evolutionary frameworks. In particular, using simulation, we can explore the effects of biased transmission in situations of cultural and biological evolution which are more difficult to produce in the lab. Xu et al.'s experiment has several practical limitations—transmission is only between single individuals, learning must happen very rapidly, and there is no feedback on production—but simulations are not so constrained. Kirby and Hurford's Iterated Learning Model, for instance, includes feedback, and can be extended to learning from more than one individual. This allows us to make more specific claims about the role biased transmission may or may not have played in actual language evolution.

These examples make clear how simulation can extend the inferential implication of empirical evidence in both ways. Simulation connects the dots between Regier et al.'s observations and particular hypotheses. Similarly, simulations can situate the mechanism demonstrated in the lab by Xu et al. in more realistic evolutionary frameworks. Simulations can thus extend the inferential reach of evidence from relatively weak sources like comparative linguistics and contemporary cognitive science. Both examples, of course, are oversimplified for expository purposes. As Steven Peck observes for ecology, the epistemology of agent-based modeling requires a "'hermeneutic circle,' a back and forth in active communication among both modelers and ecologists" (2008). The same is true in evolutionary linguistics. The role of simulation is not merely to draw out the ramifications of empirical evidence, but also to inspire new questions and hypotheses for empirical research. Nevertheless, insofar as evolutionary linguistics is going to make progress in confirming particular hypotheses, simulations play a key role in taking weak evidence and extracting as much out of it as possible.

In addition to allowing us to draw out the inferential import of evidence, simulation might also contribute to evolutionary linguistics by providing data points in a robustness analysis. In a robustness analysis, we take a claim as well-supported if it is supported by a variety of models and experiments, each of which makes different assumptions and idealizations. Irvine et al. (2013) argue convincingly that we can include simulations in robustness analyses in evolutionary linguistics. They give as an example an analysis of models of the emergence of compositionality. In a number of varied models, compositionality generally emerges only when there is a tradeoff between the learnability and expressive power of a protolanguage. That this result is consistent across models making competing assumptions, they argue, is reason to think it more likely that such a tradeoff was in fact a driving force in the emergence of compositionality.

Robustness analyses of this sort are distinct and complementary use of simulations to using them to determine the inferential import of evidence. In fact, in connecting data to theory or situating it in an evolutionary framework, we may want to use models with a variety of parameters and idealizations to get an idea of how robust the conclusions we're drawing are. But this is not to say that the other two uses of simulation in confirmation are merely varieties of robustness analysis. We won't always want or need to use robustness analysis to make use of simulations in the other two ways, and the way in which they make an epistemic contribution is by helping us understand the meaning of our data, whereas the epistemic contribution of a robustness analysis is to show us the extent to which our conclusions rest on particular assumptions and idealizations. The epistemic contributions of simulation to evolutionary linguistics thus take multiple useful forms.

Responses to the skeptics

Armed with an account of the epistemology of simulation in evolutionary linguistics, we can now address simulation skeptics.

Templeton argues against the utility of simulation in studying human evolution because “the relationships observed for the aggregate data” don’t prove “that an underlying process model is true” (2007: 1515). This objection misses the mark on two fronts. He’s correct that the following sort of argument is unsound, because premise (3) is false:

(1) The actual evolutionary endpoint of human evolution has been *x*

(2) Simulation *y* ends up approximating *x*

(3) If a simulation and an empirical process have a similar output, then the causal mechanisms of the simulation are similar to the causal mechanisms of the empirical process.

(4) Therefore, the causal mechanisms underlying the human evolution of *x* include the mechanisms found in simulation *y*

At this point, however, it should be clear that evolutionary linguists are not doomed to this sort of reasoning. First, while the argument above is unsound, the problem is that it construes simulation in terms of providing *proof*, when simulation actually draws out the import of *evidence*. It’s reasonable for a well-designed simulation with appropriate connection to empirical evidence to increase our confidence in a hypothesis about the causal mechanisms underlying human evolution, without necessarily demonstrating its truth. Moreover, researchers into human evolution don’t have to rest the entire weight of their reasoning with simulations on matching the simulation output to observed facts. As we’ve seen, simulations can connect to empirical facts in the properties and parameters they assign to elements of the simulation, and this leads to a richer sort of inference than Templeton’s target. For these two reasons, Templeton’s objection, while it may tell against some arguments made on the basis of simulation, provide no reason to be skeptical of the utility of evolutionary linguistic simulations in general.

Roberts worries that simulations often “give conflicting results, partly to be explained by the necessary simplicity of the agents that interact in them” (2010: 140). This simplicity, however, is not a pure drawback but reflects a tradeoff between realism and control. As we saw in the example of biased transmission, the simplifications in ABMs and other simulations allow us to control for variables we can’t control for in the lab, facilitate understanding the explanations produced in response to experimental work, and explore counterfactual scenarios which illuminate the import of what goes on in the lab. Simulation and empirical work each have their pros and cons, and using them in tandem (as must be the case according to the epistemology of simulation we’ve arrived at) allows us to work around these tradeoffs. Simplicity, then, cuts both ways, and its drawbacks are offset if modeling is tied to experiment and observation.

The most antagonistic skepticism to simulation in evolutionary linguistics is also the easiest to address. Hauser et al. assert that “all modeling attempts have made unfounded assumptions and have provided no empirical tests, thus leaving any insights into language's origins unverifiable” (2014: 1), but with our epistemology of simulation in hand we see that it is this assertion which is unfounded. As we’ve seen, researchers modeling language evolution aren’t just pulling evidence out of a hat. The feedback cycle between simulation and empirical evidence provides foundation for some modeling assumptions, as when an agent-based modeler draws on empirical work in designing the simulated agents. Other assumptions do lack empirical foundation, but those are often part of the hypothesis in question. So while it’s true that there are not generally knockdown empirical tests of simulation results, there are empirically-informed reasons to favor certain models. We can, and do, find some simulations insufficient when they fail to meet empirical tests, as in the toy example above of basic Lewis-Skyrms games and optimal semantic partitioning. Hauser et al. are thus mistaken about the lack of an empirical basis for inferences drawn from simulating the evolution of language.

Simulation thus appears to be a useful tool despite the skeptics' arguments to the contrary. This is not to say that we should be particularly trusting of claims made on the basis of simulated human evolution. Although I've framed this discussion of the epistemology of simulation as a response to the skeptics, one of its implications is that to have inferential power, a simulation of language evolution must have appropriate connections to empirical work. Insofar as this standard isn't always met, a healthy degree of skepticism towards claims based on simulations is warranted. Nevertheless, dismissive skepticism of the sort espoused by Hauser et al. fails to grasp the scientific potential of simulating the evolution of language.

Conclusion

I have argued that by extending the inferential reach of evidence, computer simulation can play an important role in the confirmation of hypotheses in evolutionary linguistics. My primary aim has been to justify the use of simulation in that field, but our discussion has broader implications. It provides a partial vindication not only of simulation, but also of the field of evolutionary linguistics as a worthwhile pursuit, since it shows how we can make some inferences even on the basis of the generally weak evidence available to us. Additionally, while we have focused on evolutionary linguistics, our discussion might apply other fields studying the evolution of human intangibles, such as the evolution of morality or the evolution of religion. These fields suffer from a similar lack of strong evidence, and so can similarly benefit from how simulation can extend the reach of that evidence.

References

Bickerton D (2008) But how did protolanguage actually start? *Interaction Studies*, 9(1), 169–176.

Bolhuis JJ, Tattersall I, Chomsky N, & Berwick RC (2014) How could language have evolved? *PLoS biology*, 12(8).

Cangelosi A, & Parisi D (2002) Simulating the evolution of language (Vol. 1). London: Springer.

Fitch WT, Hauser MD, & Chomsky N (2005). The evolution of the language faculty: clarifications and implications. *Cognition*, 97(2), 179-210.

Franklin A (1986) *The neglect of experiment*. Cambridge University Press.

Gong T (2011) Simulating the coevolution of compositionality and word order regularity. *Interaction studies*. 12(1), 63-106

Gong T, Shuai L, Zhang M (2014) Modeling language evolution: Examples and predictions. *Physics of Life Reviews* 11, 280-302.

Harrison SP (2003) On the Limits of the Comparative Method. *The handbook of historical linguistics*, 213. Wiley

Hauser MD, Yang C, Berwick RC, Tattersall I, Ryan MJ, Watumull J, Chomsky N and Lewontin RC (2014) The mystery of language evolution. *Front. Psychol*. 5(401).

Irvine L, Roberts SG, & Kirby S (2013). A robustness approach to theory building: A case study of language evolution. In the *35th Annual Meeting of the Cognitive Science Society* (CogSci 2013) (pp. 2614-2619).

Jackendoff R (1999) Possible stages in the evolution of the language capacity. *Trends in cognitive sciences,* 3(7), 272-279.

Jäger G (2007) Evolutionary game theory and typology: A case study. *Language*, 74-109.

Kirby S, Dowman M, & Griffiths TL (2007) Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, 104(12), 5241–5245.

Kirby S & Hurford JR (2002) The emergence of linguistic structure: An overview of the iterated learning model. In *Simulating the evolution of language* (pp. 121-147). Springer

Lekvam T, Gambäck B, & Bungum L (2014, April) Agent-based modeling of language evolution.

In *Proc. of 5th Workshop on Cognitive Aspects of Computational Language Learning* (CogACLL)@ EACL (pp. 49-54).

Martins M, Raju A, & Ravignani A. (2014). Evaluating the role of quantitative modelling in language evolution*. The past, present and future of language evolution research,* 84-93. EvoLang 9 Organizing Committee

Parker W (2008) Franklin, Holmes, and the epistemology of computer simulation*. International Studies in the Philosophy of Science* 22(2), 165-183.

Peck SL (2008) The hermeneutics of ecological simulation. *Biology & Philosophy*, 23(3), 383-402.

Railsback SF, & Grimm V (2011) *Agent-based and individual-based modeling: a practical introduction.* Princeton University Press.

Regier T, Kemp C, and Kay P (2015). Word meanings across languages support efficient communication. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence*. Wiley.

Robert T (2010) Saussure et l'origine du langage: un interdit à dépasser par la philosophie linguistique. *Rivista Italiana di Filosofia del Linguaggio*, (3), 147-156.

Roberts G (2010) An experimental study of social selection and frequency of interaction in linguistic diversity*. Interaction Studies*, 11(1), 138-159.

Skyrms B (2010) *Signals: Evolution, learning, and information*. Oxford University Press.

Templeton AR (2007) Genetics and recent human evolution. *Evolution* 61(7) 1507-1519.

Tomasello M (2002). Some facts about primate (including human) communication and social learning. In *Simulating the evolution of language* (pp. 327-340). Springer London.

Winsberg E (2009) Computer simulation and the philosophy of science. *Philosophy Compass,* 4(5), 835-845.

Wray A (2002) *Formulaic language and the lexicon.* Cambridge University Press.

Xu J, Dowman M, & Griffiths TL (2013) Cultural transmission results in convergence towards colour term universals. *Proceedings of the Royal Society B: Biological Sciences*, 280(1758).

CHAPTER 5: THEORY OF MIND AND COMMUNICATIVE INTENTIONS IN SCIENTIFIC MODELS OF LANGUAGE USE

I. Introduction

Following Grice's seminal "Meaning" (1957), a diverse array of theories in philosophy, psychology, and linguistics have explained linguistic meaning and communication in part by appeal to interlocutors' higher-order communicative intentions. Call this feature of these theories *intentionalism*:

> INTENTIONALISM: A theory is intentionalist if it claims linguistic communication depends on
>
> (a) the speaker's (writer's, signer's) intention to achieve some effect by means of a change in an addressee's intentional state
>
> and
>
> (b) the addressee recovering meaning by reasoning about the speaker's communicative intentions

Under this characterization, intentionalists include not only Grice[44] (1957) and later Griceans (Schiffer 1972, Levinson 1983, Horn 1984), but also Relevance Theory[45] (Sperber and Wilson 1995), common ground theories[46] in psycholinguistics (Clark 1996, Tomasello 2008) and philosophy (Stalnaker 2002), game-theoretic treatments of meaning[47] (Lewis 1969, Parikh 2000, Clark 2011, Jäger 2012), speech act theories[48] (Austin 1975; Searle 1969), and politeness

[44] For whom both non-natural meaning and indirect speech acts involve higher-order intentions.

[45] Which involves interlocutors considering each other's desired information and beliefs about the utterance context.

[46] In which language use depends on conversational participants' beliefs about what knowledge they share with their interlocutors.

[47] Which models how agents make decisions based on their beliefs about how other agents will make decisions.

[48] Which see one component of meaning as the effect of an utterance on the intentional state of the addressee.

theory[49] (Brown and Levinson 1987), among other examples. While these accounts differ widely in how they explain linguistic communication, they all share a central commitment to the claim that communication depends on the exercise of *theory of mind,* the ability to reason about the mental states of other creatures.

This leaves an otherwise disparate set of theories vulnerable to the same objection:

> THEORY OF MIND INESSENTIAL (TOMI): there are frequent, non-trivial cases of successful linguistic communication where one or more interlocutors do not exercise theory of mind to produce higher-order communicative intentions.

Although controversial, TOMI has the support of multiple lines of empirical research, some of which I will survey in section II. I am less concerned with establishing the truth of TOMI, however, than with determining the intentionalist's best response. Intentionalist theories include some of the most important accounts of linguistic communication from several disciplines, and we should not give those theories up too quickly.

A number of extant responses to TOMI attempt to reinterpret the data behind the empirical objection in ways more friendly to intentionalism. I will show that the most promising version of this tactic falls short of showing TOMI to be false. We then consider whether the intentionalist can avoid TOMI by loosening the intentionalist criteria. I will argue that this gives up too much, because it abandons the real insights found in appealing to higher-order intentions in communication. A third approach is to limit the aims of intentionalist theories. Most commonly, some philosophers have argued that intentionalism in meant not as a descriptive account of human psychology, but merely as a rational reconstruction, which does not need to answer to the facts in the same way that descriptive theories do. This response also abandons too much. Many intentionalists do have descriptive aims, and it is implausible to see them as presenting rational reconstructions.

The rational reconstruction approach is on the right track, however. Drawing on some of

[49] Which has a Gricean core.

Grice's comments, I will argue that intentionalists are best construed as presenting idealized scientific models. As with a rational reconstruction, the validity of a scientific model is consistent with its falsity. But unlike rational reconstructions, models can aim at accurately representing their target system. Construing intentionalism as a modeling strategy thus allows us to both acknowledge TOMI and preserve intentionalism as a scientifically valid account of linguistic communication.

II. Evidence that theory of mind is inessential to linguistic communication

TOMI has a wide variety of empirical support, and I have space to discuss only some of the more striking examples of research showing that linguistic meaning and communication does not depend on the intentionalist criteria. We will look in detail at three flavors of this objection: linguistic individuals without full theory of mind, communication based on egocentric reasoning, and reflexive, non-intentional communication.

Children and individuals with non-linguistic impairments

The first variety of the objection points to a class of people who successfully communicate, but do not seem to be able to have intentions regarding the mental states of their interlocutors. Breheny (2006), for instance, argues that the case of young children is damning for intentionalist theories, specifically targeting Grice, Lewis, Stalnaker, and Relevance Theory. The argument runs as follows:

(1) 3-year-olds are competent language users (by developmental linguistic standards).

(2) 3-year-olds lack theory of mind (according to developmental psychology).

(3) Therefore, theory of mind is not required for linguistic competence.

Since the intentions at the heart of intentionalism involve theory of mind, intentionalism must be false, argues Breheny.

The evidence for (1) is that by age three most children, while not as fluent as adult language-users, frequently produce utterances identical to the utterances an adult would employ in the same circumstances. The evidence for (2) largely comes from the false-belief task (Wimmer and Perner 1983). In the false-belief task subjects are presented with a doll named

Sally, who caches a candy in a cupboard. Sally leaves and Anne enters, relocating the candy to a box. Anne leaves, Sally returns, then the subject is asked where Sally will look for the candy first. The correct answer, of course, is the cupboard, but subjects under the age of four consistently fail the task, saying Sally believes it to be in the box. Meta-analysis (Wellman, Cross, and Watson 2001) shows this result to be robust, occurring in most replications and insensitive to whether Sally and Anne are portrayed by dolls or human actors. This is taken to be evidence that young children either do not have or do not employ theory of mind, and instead use egocentric techniques (i.e. asking themselves "What do I know? rather than "What does Sally know?") to make predictions about other agents. Thus there is robust[50] evidence that children under four use language without reasoning about their interlocutors' mental states as intentionalism requires.

A second set of cases used to make the same point are cases of linguistically-competent individuals with an impairment that may render them unable to exercise the requisite theory of mind. Laurence (1996) appeals to these sorts of cases as counterexamples to Lewis' game-theoretic account of language. The argument is parallel to the argument in the case of young children. Individuals with moderate autism spectrum disorder[51] are often able to communicate linguistically, but consistently fail the false belief task and so purportedly do not have strong theory of mind capacities. In this case as well as the case of young children, we have individuals who seemingly both produce and interpret meaningful utterances without engaging in reasoning about their interlocutors' mental states. Therefore, their language use cannot depend on higher-order communicative intentions, and intentionalism appears to be false.

Egocentric reasoning in everyday language use

A second variety of empirical evidence against intentionalism involves claims that the intentionalist criteria are not met even in many cases of everyday, unimpaired, adult linguistic

[50] Though not universally accepted (see section III.)

[51] Laurence uses a different example—Williams syndrome—but individuals with Williams syndrome are probably not a good counterexample to intentionalism (Tomasello 1995). Autism spectrum disorder better fills that role in Laurence's argument.

behavior. Peters (2015), for instance, raises cases of this sort as evidence against Tomasello. If communication occurs in line with the intentionalist paradigm, then speakers should take into account the mental states of their audience, and vice versa. Some clever experiments, however, seem to show that this is not the case. Keysar et al. (2003) gave subjects the simple task of following instructions given to them by a director, a covert confederate of the researchers. An array of shelves stood between the subject and the director. Some shelves were occluded on one side, so that the subject but not the director could see the contents of that shelf. This meant that reference could differ between the director's perspective and the subject's perspective. For example, if the shelves held three candles in three different sizes, but the smallest was on an occluded shelf, the smallest candle visible to the subject would differ from the smallest candle visible to the director (see figure 1). Thus, if the subject were employing theory of mind to interpret the director, she would take the director to mean by "the small candle" the second smallest candle. If, however, the subject was using egocentric methods to interpret the utterance, she would interpret “the small candle” to mean the smallest candle, despite the fact that she should know that the director is unaware of its existence.

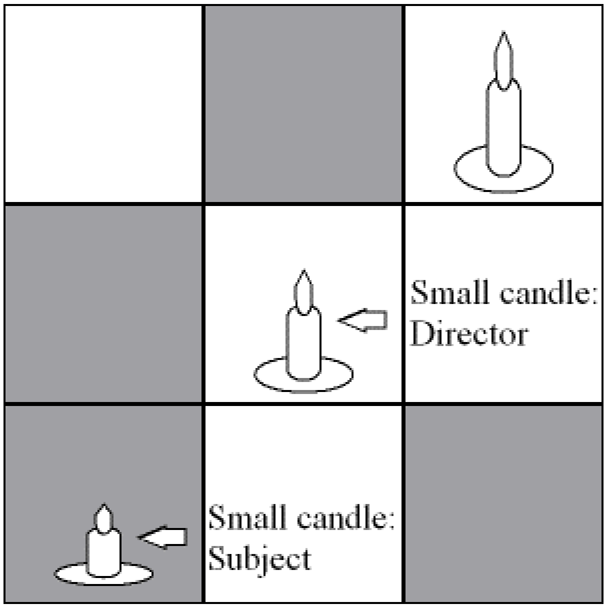


Figure 1: Gray boxes represent shelves occluded from the director’s view.

In conditions with indeterminate reference such as the example above, 71% of subjects tried to move the occluded object at least once, and 46% of subjects did so in more than half of the experimental trials[52]. Variants on this paradigm involving different sorts of ambiguity, placing

[52] Keysar et al. were careful to rule out alternative explanations of this failure. Subjects never committed the same sort of error in a baseline condition, demonstrating that the error must be due to interpretive failure, rather than some other cause such as motor error. Additionally, in follow-up interviews only one

the subjects in the role of director, more naturalistic and free-form tasks, and quantitative eye-tracking and neurophysiological measures have reliably extended these results (Horton and Keysar 1996, Keysar et al. 2003, Lane et al. 2006, Apperly et al. 2010, Savitsky et al. 2011).[53] Keysar and colleagues have argued on this basis that both speakers and addressees frequently fail to consider the intentional attitudes of their interlocutors during communication. If they are right, this seems to be additional evidence that intentional reasoning is unnecessary to communication. The claim is not that subjects lack theory of mind abilities, nor that they never use them in communication, only that sometimes, perhaps most of the time, they do not. So long as this occurs with non-trivial frequency, then everyday communication does not typically depend on higher-order communicative intentions and intentionalism is false.

Uttering without intending

The objections to intentionalism we have considered thus far claim that linguistic communication frequently occurs in the absence of one of the intentionalist criteria. Either individuals can communicate without theory of mind, or even competent language users often resort to egocentric rather than intentionalist methods in production and interpretation. The final objection we will consider goes even further, denying that linguistic communication need involve intentions at all.

Lind et al. (2014a; 2014b) argue that speech is often reflexive, not intentional. They provide evidence that our articulatory behavior often occurs in the absence of conscious intention to mean what we say, and the phenomenology of intention—feeling like we intended to say what we said for some reason or another—is the result of reflection on our utterances after the fact. In their Stroop task, subjects were required to produce linguistic responses to the color of a presented stimulus. Stimuli were printed in colors other than the color denoted by the printed word, and subjects were instructed to name the color of the text. For example, a subject might

out of thirty-eight subjects reported suspecting that the director was a confederate, so subjects would not have known that the director was actually aware of the hidden object.

[53] See Brown-Schmidt et al. (2003) for partial dissension.

see the word ‘blue’ displayed in red text, in which case the subject should say “red.”

In Lind et al.’s experiment the subject performs the task while wearing headphones. In the experimental condition, the headphones play a recording of one of the subject’s previous answers simultaneous to their present response. “Green,” for instance, might play on the headphones as the subject articulates “gray.”' When the onset of the headphone output coincided closely with the onset of the subject’s utterance, only 32% of subjects suspected some sort of manipulation, and only 4% were certain that the researchers had played a replacement word over the headphones. Of the remaining 68% of subjects, 38.5%, when asked, claimed that they had uttered the replacement word. An additional 16.5% did not need to be asked, because they corrected themselves, saying something like “no, green” after hearing “gray,” even though they had uttered “green” in the first place. An additional 29.7%, when asked, did not initially claim to have uttered the replacement word, but revised this answer upon reflection, ultimately admitting to have uttered the word they did not actually utter. Summing these up, 85% of these subjects believed they had said something they had not actually said.

In their analysis, Lind et al. (2014a: 8) suggest that the explanation for these results is that "we actively use feedback to help specify for ourselves the full meaning of what we are saying. In effect, we propose that auditory feedback provides us with a channel for high-level, semantic 'self-comprehension.'" To put it more bluntly, sometimes our perception that we intended to mean *p* by an utterance is an illusion. We use cues, including hearing our own voice, to determine what we said, and infer that we must have intended to mean *p* because the effect of what we said was to mean *p*. In reality, however, our linguistic behavior was reflexive, occurring in the absence of such a linguistic intention. Because Lind et al.’s experiment focuses on a common linguistic task—naming—it suggests that reflexive linguistic behavior is common. That their results are more than a fluke is suggested by the similar outcome in Banakou and Slater (2014), who use virtual reality to fool subjects into thinking they both uttered and intended to utter a word actually uttered only by their virtual avatar. Given these experiments it seems that there are cases of meaning which are not the product of conscious intention. Although these experiments fall short of proving that that linguistic behavior is frequently reflexive rather than intentional, since the relevant intentions may be opaque even to their bearer, it does provide telling evidence to that effect.

Furthermore, these experiments explain why intentionalism is so intuitive. Subjects have the phenomenology of intention even when the subjective experience of intending does not match their actual behavior. Consequently, when we as philosophers or linguists reflect on our own linguistic behavior, our intuitions, guided by the phenomenology, will tend towards intentionalism. But since the phenomenology can be misleading, our intuitions are unreliable, undermining any armchair evidence in favor of intentionalism. If we really do utter without intending, as Lind et al. argue, intentionalism is in trouble.

III. Responses to TOMI

Denying the empirical inference

As the examples above demonstrate, multiple lines of empirical evidence seem to support TOMI. One avenue of response for the intentionalist is to produce an alternative explanation for the problematic empirical observations. The most plausible of these responses appeals to failures in executive function (EF), the mental capacity to regulate, control, and inhibit other cognitive functions. Some evidence (Bradford et al. 2015, Ferguson et al. 2015) seems to indicate that taking the perspective of others is particularly demanding on EF, perhaps because it requires inhibiting egocentric processes. If this is so, apparent theory of mind failures are consistent with the presence of theory-of-mind-laden processing, since we can attribute egocentric behavior to EF failure instead. Thompson (2014), for instance, defends intentionalism in the face of the false-belief task by attributing children's failure to a deficit in EF, not theory of mind, citing Onishi and Baillargeon (2005) as evidence. The idea is that young children are aware of Sally's false belief, but lack the mental control to keep themselves from sharing their own knowledge instead. Similarly, the intense world hypothesis (Markram, Rinaldi, and Markram 2007), denies that autistic individuals lack theory of mind, and gives a similar EF explanation for why they fail the false-belief task. Following the lead of these responses, we could argue that

Keysar's experimental results are due not to absence of intentional reasoning, but instead to failure to inhibit egocentric cognitive processes. If this line of response is correct, then higher order intentions might always be present in linguistic communication, but they sometimes fail to win out in competition with other cognitive determinants of behavior.

These alternative explanations for evidence purporting to undermine intentionalism are worth pursuing, but the orthodox interpretations of the false-belief experiment, Keysar's work, and so on tell against intentionalism. Even if the alternative explanation turns out to be correct in some cases, it seems unlikely that this will be so in every case. Additionally, even if children, individuals with autism, and so forth do have some theory of mind[54], it seems incontrovertible that they are unable to exercise it fully. This line of response is thus predicated on a shaky empirical bet.

Moreover, even if intentional reasoning is always present but sometimes suppressed, mere presence is not enough to guarantee the truth of intentionalism. Intentionalism is meant as an explanation of linguistic communication, and in cases where theory of mind is suppressed, the actual explanation for the communicative behavior will rest on the egocentric processes which won out. Arguments from failures of executive function thus fail to undermine TOMI, and the intentionalist must look for a stronger line of response.

Tweaking the intentionalist criteria

Since dismissing the empirical data is unlikely to succeed, a defender of intentionalism might have to bite the bullet, granting the empirical cases. This bullet-biting can take two forms: we can weaken the intentionalist criteria so as to recover the previously excluded cases, or we can allow that many cases of linguistic communication fail to meet our criteria. An example of the first route is Moore (2013), who argues that Lewisian conventions can get off the ground even if

[54] Young children, even infants, *do* seem capable of attributing agency of some sort to external beings, as shown in experiments discussed in Carey (2009) Ch. 5. The ability to exercise theory of mind, however, involves being capable of deploying attributions of intention in various sorts of reasoning, not mere recognition of intentional states. The false belief task and similar experiments seem to show that young children lack the ability to fully exercise theory of mind in this sense.

we substitute ostention and imitation for some of the higher-order mental representations required in Lewis' original theory. Moves like Moore's successfully escape the empirical objection, but I doubt that we can really characterize them as satisfactory responses for the intentionalist. Yes, we can take Lewisian or neo-Gricean theories and swap out the intentionalist criteria for substitutes such as imitation and joint attention, but the resulting theories are not intentionalist at all, because they no longer implicate theory of mind in linguistic communication. The explanatory power of intentionalism comes from the sophisticated reasoning it attributes to linguistic agents, and much of this power is lost if we remove this sophistication. To bite the bullet in this way is to give up the game.

Rational reconstruction

A better option for the intentionalist is hold to the strict definition of intentionalism but relax its empirical scope. In one common form, this approach argues that intentionalism is not a claim of descriptive psychology, but instead a rational reconstruction. Rational reconstructions are not meant to be predictive, or to describe actual causal mechanisms, but instead are meant make sense of a phenomenon by interpreting it as rational activity. Philosophers of language sometimes argue that objections like TOMI miss the mark because intentionalism is merely a rational reconstruction. “Semantic and pragmatic theories are rational reconstructions of the ability of speaker-hearers to interpret uses of sentences,” claims Soames, “The cognitive processes by which this occurs are not our concern” (2010: 171-72). Likewise, Bach argues that “Grice did not intend his account of how implicatures are recognised as a psychological theory nor even as a cognitive model. He intended it as a rational reconstruction...He was not foolishly engaged in psychological speculation about the nature of or even the temporal sequence of the cognitive processes that implements that logic” (2006: 25, see also Saul 2002, O’Rourke 2003). For reasons I’ll discuss below, I’m not convinced that this is a fair representation of Grice. It is clear however, that some intentionalist philosophers can escape the force of TOMI because the aim of their theorizing—rational reconstruction—does not require empirical accuracy.

As a general response to TOMI, however, the retreat to rational reconstruction is inadequate. Even if it is true that intentionalist philosophers of language only intend to produce rational reconstructions[55], many intentionalists are scientific psychologists and linguistics. These psychologists and linguists draw on the work of philosophers of language (especially Grice and Austin) in crafting scientific theories. Scientific theories, unlike rational reconstructions, generally aim for explanation in terms of actual causes[56], in this case an account of the actual psychology underlying linguistic communication. Borg has articulated precisely this problem for rational reconstructions of linguistic meaning (2009: 34):

> My worry is that talk of 'rational reconstruction' runs the risk of driving too great a wedge between the semantic theory and the psychological theory, for if all one is offering is a way in which speaker meaning could be recovered, with no requirement that ordinary speakers do recover meaning in this way, then we seem to be sliding away from a picture which treats semantic content as dependent on psychological content and towards an account which treats semantic content and psychological content as more or less independent of each other

Empirically inclined intentionalists thus can't evade TOMI by retreat to rational reconstruction. Nevertheless, the retreat to rational reconstruction prefigures a tactic that will work for intentionalists in general, one in the spirit if not the letter of Grice.

IV. Intentionalism as idealized modeling

In *Aspects of Reason*, Grice characterizes his inquiry into reasoning as an act of creating "models or ideal constructions" which can help us "understand actual reasonings" (2001: 7-8). If we extend the scope of these remarks to Grice's work on meaning, it gives us the beginning of a general response to TOMI. Grice takes his philosophizing to be model building in a broad sense,

[55] Some philosophers who are intentionalists clearly aim for more than rational reconstruction, but arguing over who falls into this category doesn't concern us here.

[56] We might also want to use rational reconstruction to produce an account of ordinary language users' *post hoc* interpretations of their own language use, perhaps because such interpretations play a role in holding each other to account for "what is said" (see Camp 2006). This aim, however, is mostly distinct from explaining primary linguistic cognition.

including the construction of “analytic models,” which are meant to aid in philosophical categorization, and “normative models,” which provide examples for agents to aspire to (2001: 8).[57] Neither of these types of models will rescue intentionalism as a descriptive psychological theory. But Grice also takes himself to be putting forward “explanatory models” which “play a central part in providing” an explanation of the phenomenon in question (2001: 8). In the case of reasoning, for instance, a model of reasoning drawn from first order logic might not accurately describe normal human reasoning processes, but it does explain why the actual processes make successful inferences (2001:9-10). If we apply the same sort of treatment to intentionalism, this leads us to see intentionalist theories not as direct descriptions of the psychological processes necessarily involved in linguistic communication, but as *idealized scientific models* of communication.

Scientific models can be employed for a variety of purposes, including prediction, but even predictive models are not meant to perfectly represent or simulate their target systems (Weisberg 2013). Instead, when scientists create a model, "the real systems in their respective domains of inquiry are knowingly and systematically misrepresented" (Jones 2005: 2). These misrepresentations, or *idealizations*, serve a number of purposes, including making it easier to analyze models and highlighting particular attributes of systems of interest. Idealization occurs in every science. Physicists construct mathematical models which omit the effects of friction or the gravitational forces exerted by small objects; biologists sometimes ignore the fact that populations are composed of finite numbers of discrete individuals; economists sometimes pretend that consumers are better informed than they actually are. These intentional falsehoods make complex problems tractable, which is why models can useful even if they are not strictly true.

Such idealized models are central to scientific practice, and intentionalists can avoid the empirical objections by construing their positions as putting forward scientific models of communication. Like Grice, we can see intentionalism as putting forward an ideal case of

[57] He makes similar, but less explicit, remarks about meaning in “Meaning Revisited” (Grice 1991).

communication. Instead of Grice's *normatively* ideal case, however, scientific intentionalists focus on a *descriptively* ideal (idealized) case—a model. Because models are idealized, they are necessarily false. The mere fact that a model inaccurately describes a number of empirical cases does not, on its own, constitute a reason to reject the model. Whether a model is a good model is a pragmatic question: does the model improve our understanding or increase our predictive or explanatory abilities? An idealization is licensed if it serves these ends, and unwarranted if it does not. To show that we can save intentionalism by construing it as a modeling strategy, then, I need to demonstrate that the empirical exceptions to intentionalism are the product of justifiable idealizations. To establish this, I will outline how intentionalist models of communication work as *special case models* and *minimal models*.

Special case models

A number of scientific fields use models of special cases, such as optimal or limiting cases, to great success. Examples include evolutionary optimality modeling, rational choice theory, and ideal observer analysis. The idea behind this sort of modeling is that we can "describe one class of cases, which are simple and tractable, and use these as the basis for a more indirect understanding of the others" (Godfrey-Smith 2009: 4). Cases of communication which depend on higher-order communicative intentions are just such an illuminating special case in a number of ways.

First of all, for predictive purposes, that the actual psychology of communication does not always meet the intentionalist criteria is of secondary concern. As famously pointed out by Dennett, in several domains of science we find it helpful in making predictions to ascribe intentions even where we suspect none exist. This is true in biology when, in order to reconstruct evolutionary history, we pretend that organisms were designed with intent (Dennett 1996). And it is true in psychology, when we use folk-psychological ascriptions to help us predict others' behavior and to bootstrap our understanding of mental processes (Dennett 1989). So it should come as no surprise that intentionalist models can serve as useful predictors of linguistic behavior. Modeling the special case of intentional communication allows us to make predictions about linguistic behavior in general.

Of course, if our requirement for the application of the intentional stance is mere predictive success, it overgenerates. We can make successful predictions in almost any domain by ascribing intentions—we can pretend that positively charged particles *desire* to attract negatively charged particles, and thus make successful predictions about electromagnetism, for instance. But scientists do not find it useful to create predictive intentional models in particle physics. Why then, is the intentional stance useful in biology and psychology? Because, unlike the case of particle physics, the intentions we ascribe are motivated by a parallel between biological needs and the motivations of a hypothetical agent. Evolution acts as if it had intentions because the adaptive pressures driving evolution are the same adaptive pressures that an intelligent designer would engineer organisms to meet. Sub-rational psychological process act as though they worked on an intentional level because they are responding to the same organismal needs that intentional reasoning does. The intentions we might ascribe to positively charged particles, on the other hand, are ad hoc. We can create predictively successful intentional models of particle physics only after we already understand the behavior of the relevant physical systems, because there is nothing in their behavior that responds to circumstances the way an intentional agent would. So application of the intentional stance for predictive purposes is licensed only when there is a motivated parallel between the system in question and how an intentional agent would act.

This criterion is clearly met in the case of intentional models of meaning. The communicative goals that linguistic organisms respond to—coordination, information transfer, deception, etc.—are equivalent whether or not the organism is employing theory of mind to meet those needs. Behavior in the special case, where higher-order communicative intentions are in play, will thus generally resemble behavior in other cases. For predictive purposes, then, the idealization to the special case of intentional communication is warranted. It will not, of course, yield perfect predictions, but neither do special case models in other domains. Such models remain in use because some predictive error is tolerable, particularly if offset by other factors such as ease of use and understandability.

As special case models, intentionalist models also contribute to explanation. To see how, consider how optimality models are used evolutionary biology. A typical case of optimality modeling in behavioral ecology runs as follows: we consider a number of factors contributing to an organism’s fitness, measure all but one, then, assuming that the organism is maximally fit given its environment, use the measured quantities to estimate the unknown quantity. For example, we might look at available food sources to estimate the average jaw size of a predatory species. Or we might estimate the danger per minute a bird exposes its offspring to in leaving the nest unattended in conjunction with a measure of how many calories it is able to gather per minute to infer the amount of time it spends foraging for food for its chicks.

As descriptive accounts of behavior and evolution, however, optimality models are open to Gould and Lewontin’s (1979; see also Orzack and Sober 1994) famous objection that these models falsely assume that organisms are always maximally adapted to their environment. Gould and Lewontin's objection falls short for the same reason TOMI falls short. They fail to take into account the fact that optimality models are idealized by design, and can be evaluated only in respect to particular modeling goals. As Parker and Maynard Smith argue in response to Gould and Lewontin, optimality models "serve to improve our understanding about adaptations rather than to demonstrate that natural selection produces optimal solutions" (1990: 27; see also Potochnik 2009). Even when optimality models fail predictively, they help us understand which causal components of an evolutionary process contributed in any particular case—if optimality obtains, natural selection has almost certainly dominated, but if not, we need to look for other evolutionary forces as well. Moreover, by providing a limiting case optimality models give us a baseline from which we can determine the degree of contribution of all sorts of evolutionary forces, not merely the adaptive ones. As a particularly tractable special case, optimality models sort out these contributing factors and thus contribute to scientific explanation.

Similar useful features explain the persistence of ideal-agent rational choice modeling long after researchers (Simon 1955) called attention to their empirical inadequacy. To give one example, when looking for rationality-approximating heuristics the rational choice model still gives researchers a baseline to compare with the heuristic's performance. See, for instance, the methodology[58] used by Czerlinski, Gigerenzer, and Goldstein (1999), where they compare the predictive success of hypothesized psychological heuristics against the performance of statistically-sophisticated ideal rational agents. By showing that the Take the Best heuristic performs nearly as well as multiple regression models, they show why human psychology would employ the Take the Best: it performs nearly as well as is possible, but only requires limited resources. And by answering that 'why' question, they provide an explanation for the phenomenon. Human reasoners only employ multiple regression in decision making on rare occasions. Nevertheless, a model of an agent in this special case helps us understand why actual decision makers use the simple heuristics they generally do. Additionally, it sets a baseline for maximum possible success that helps us understand why failures of reasoning can be considered failures. Thus, this special case model of an ideal rational agent is explanatory in the same way that Grice takes his account of an ideal reasoner to be explanatory in *Aspects* (2001:9-10).

Intentionalism sometimes functions as an explanatory special case model in the same way as optimality and rational choice models. For example, consider Lee and Pinker's (2010) game-theoretic analysis of indirect speech acts. They present a scenario in which a speeding motorist, pulled over by police, attempts to offer a bribe in place of a ticket. The explanandum is that the motorist offers the bribe indirectly, saying "So maybe the best thing to do would be to take care of that here" rather than something like "If I pay you $50, will you let me off with a warning?" Lee and Pinker give a mathematically-modeled explanation, which relies on both participants reasoning about the other's communicative intentions. It is beyond unlikely that an actual motorist and officer in such a situation would actually do all the math in Lee and Pinker's explanation. Conversely, it is rather likely that the motorist would not engage in higher-order reasoning about the officer's higher order reasoning about the motorist's communicative intention. This may be, for

[58] This is similar to the methodology used in ideal observer studies of perception, another case illustrating the same points.

instance, because the motorist has made a habit of offering bribes with that precise phrase. Despite these empirical inadequacies, however, the game-theoretic model presents an explanatorily illuminating special case. Motorists offer indirect bribes for the same reason whether or not they are engaging in mathematical calculations and higher-order intentional reasoning: plausible deniability. Moreover, the special case of the game-theoretic, intentionalist reasoner explains by way of contrast why the motorist who offers a straightforward bribe has committed a communicative anomaly. As this example illustrates, by idealizing to a special case, intentionalism can produce predictive and explanatorily successful models.

Minimal models

Intentionalist models are also sometimes *minimal models*, models which abstract away from the complexity of actual systems to focus on "core causal factors which give rise to a phenomenon" (Weisberg 2013: 100). Consider supply-demand models of pricing. Market prices are the result of a number of factors, including information asymmetries, cognitive biases, government intervention, labor issues, etc. (Barro 1994). The paradigmatic economic model idealizes away from all these and presents price as a function merely of supply and demand. Because supply and demand are frequently the core causal factors, this idealization produces a minimal model which attains explanatory and predictive success in some circumstances while maintaining simplicity and tractability.

Communication is messy. Sometimes higher-order intentions play an important role, but so do a number of other factors. Nevertheless, insofar as higher-order communicative intentions are frequently core causal factors in linguistic communication, intentionalism can yield successful minimal models of linguistic communication. Take, for example, Clark's (1996) account of language use, which models interlocutors as cooperatively keeping track of mutual knowledge, a behavior requiring higher-order reasoning about mental states. Clark's model is highly idealized. As Keysar's experiments demonstrate, speakers fail to keep perfect track of mutual knowledge. Additionally, linguistic communication is embedded in a complex set of activities, and cooperation and the common ground of mutual knowledge are not the only factors in play. They are frequently, however, core causal factors—Clark presents compelling empirical evidence that interlocutors often make linguistic decisions on their basis. Abstracting away from other complicating features of communication thus yields a relatively tractably minimal model which Clark uses to explain diverse features of language use. His model, for instance, allows him to explain the essential conversational role of non-verbal signals of understanding, such as head nods and hand gestures, as a means for interlocutors to coordinate on what is mutual knowledge. Similarly, Clark explains why linguistic behavior differs when there are more than two interlocutors by appealing to the fact that speakers in such situations recognize that they share different mutual knowledge with different participants. By focusing on a single core feature of communication—the role of higher-order intentions—Clark thus produces an explanatorily powerful minimal model.

Minimal models such as these are meant to be used in tandem with methods which do not make the same idealizations. This means that models of meaning can be productive when used in conjunction with other philosophical, psychological or neuroscientific models. On the account of intentionalism as minimal modeling developed here, one can without conflict embrace intentionalism as well as conventionalism, psychologism, or most other accounts of meaning, since they too idealize away from some features of communication to focus on a limited number of important factors. Thus, even though intentionalism is empirically inadequate, as a modeling strategy it can help us understand and explain the psychology of human communication.

V. The virtues of construing intentionalism as idealized modeling

I have argued that although the facts motivating TOMI are probably true, the objection does not undermine the value of intentionalism. Intentionalists provide idealized models of a sort similar to special case and minimal models in other disciplines. These models are strictly speaking false, but purposefully so. Their validity is determined by their usefulness, not their truth. I have aimed to show how intentionalist models of communication are useful, but the strongest evidence of their utility is their persistent ubiquity in research on linguistic communication.

Treating intentionalism as a modeling strategy is successful where other responses to TOMI fail for a number of reasons. It does not require us to place risky empirical bets that future experiment will overthrow widely-accepted counterexamples to theory of mind use in linguistic communication. Nor does it have to give up on the scope of intentionalism, since intentionalist models can be useful across all the domains intentionalists work on. Moreover, it accomplishes all this without being overly reductive. In treating intentionalism as a modeling strategy we do not take something as complex as linguistic meaning and reduce it to a single factor. Instead, we provide a tool for exploring one aspect of human communication; a tool, for that matter, which works well in conjunction with our other tools.

Perhaps the most important virtue of my response to the empirical objection is that it captures what intentionalists actually seem to be doing. They abstract and idealize away from some features and instances of linguistic cognition to focus on particularly illuminating factors and cases, and they refine the resulting models, combine them with other models, use them to guide experimental design—all key features of model-driven science. My account is even in the spirit of Grice. Recall that in *Aspects* he sees himself as creating idealized models, so in construing intentionalism as creating idealized models we not only save it from the strongest objection against it, but we recover those Gricean insights.

References

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *The Quarterly Journal of Experimental Psychology*, *63*(6), 1201-1217.

Austin, J. L. (1975). How to Do Things with Words.

Bach, K. (2006). The top 10 misconceptions about implicature. *Drawing the Boundaries of Meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn*, 21-30. John Benjamins.

Banakou, D., & Slater, M. (2014). Body ownership causes illusory self-attribution of speaking and influences subsequent real speaking. *Proceedings of the National Academy of Sciences*, 111(49), 17678-17683.

Barro, R. J. (1994). The aggregate-supply/aggregate-demand model. *Eastern Economic Journal*, 20(1), 1-6.

Borg, E. (2009) Minimal semantics and the nature of psychological evidence. In *New Waves in Philosophy of Language*, ed. S. Sawyer. Palgrave. 24-40.

Bradford, E. E., Jentzsch, I., & Gomez, J. C. (2015). From self to social cognition: Theory of Mind mechanisms and their relation to Executive Functioning. *Cognition*, 138, 21-34.

Chicago

Breheny, R. (2006). Communication and folk psychology. *Mind & Language*,*21*(1), 74-107.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language*. Cambridge University Press.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*(3), 1122-1134.

Elisabeth, C. (2006). Contextualism, metaphor, and what is said. *Mind & language*, *21*(3), 280-309.

Carey, S. (2009). *The Origin of Concepts.* Oxford University Press

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Clark, R. (2011) *Meaningful Games: Exploring Language with Game Theory.* MIT Press.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). "How good are simple heuristics?" In *Simple heuristics that make us smart*. Oxford University Press.

Dennett, D. C. (1989). *The intentional stance*. MIT press.

Dennett, D. C. (1996). *Darwin's Dangerous Idea: Evolution and the Meaning of Life*. Simon and Schuster.

Ferguson, H. J., Apperly, I., Ahmad, J., Bindemann, M., & Cane, J. (2015). Task constraints distinguish perspective inferences from perspective use during discourse interpretation in a false belief task. *Cognition*, 139, 50-70.

Godfrey-Smith, P. (2009). *Darwinian Populations and Natural Selection.* Oxford University Press.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *205*(1161), 581-598.

Grice, H. P. (1957). Meaning. *The philosophical review*, 377-388.

Grice, H. P. (1991). *Studies in the Way of Words*. Harvard University Press.

Grice, H.P. (2001). *Aspects of Reason*. Clarendon Press.

Horn, L. R. (1984). Towards a new taxonomy for pragmatic inference: Q-based and R-based implicature, in D. Schiffrin (ed.), *Georgetown University Round Table on Languages and Linguistics*, 11–42. Georgetown University Press.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground?. *Cognition*, *59*(1), 91-117.

Jäger, G. (2012). Game theory in semantics and pragmatics, in C. Maienborn, P. Portner & K. von Heusinger (eds.), *Semantics. An International Handbook of Natural Language Meaning*, Vol. 3, Berlin: de Gruyter, 2487-2516.

Jones, M. R. (2005). Idealization and abstraction: A framework. *Idealization XII: Correcting the* *model. Idealization and abstraction in the sciences*, *86*, 173-217.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32-38.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25-41.

Lane, L. W., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! Speakers' control over leaking private information during language production. *Psychological science*, 17(4), 273-277.

Laurence, S. (1996). A Chomskian Alternative to Convention-Based Semantics. *Mind*, *105*, 418.

Lee, J. J., & Pinker, S. (2010). Rationales for indirect speech: the theory of the strategic speaker. *Psychological review*, *117*(3), 785.

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014a). Auditory feedback of one’s own voice is used for high-level semantic monitoring: the “self-comprehension” hypothesis. *Frontiers in human neuroscience*, *8*.

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014b). Speakers’ Acceptance of Real-Time Speech Exchange Indicates That We Use Auditory Feedback to Specify the Meaning of What We Say. *Psychological Science*.

Levinson, S. 1983. *Pragmatics*. Cambridge University Press.

Lewis, D. (1969). Convention: A Philosophical Study.

Markram, H., Rinaldi, T., & Markram, K. (2007). The intense world syndrome–an alternative hypothesis for autism. *Frontiers in Neuroscience*, *1* (1), 77.

Moore, R. (2013). Imitation and conventional communication. *Biology & Philosophy*, *28*(3), 481-500.

O'Rourke, M. (2003). The scope argument. *The Journal of Philosophy*, 100(3), 136-157.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?. *Science*, *308*(5719), 255-258.

Orzack, S. H., & Sober, E. (1994). Optimality models and the test of adaptationism. *American*

Naturalist, 361-380.

Parikh, P. (2000). “Communication, Meaning, and Interpretation.” *Linguistics and Philosophy* 23 (2): 185–212.

Parker, G. A., & Smith, J. M. (1990). Optimality theory in evolutionary biology. *Nature*, *348*(6296), 27-33.

Peters, U. (2015). Human thinking, shared intentionality, and egocentric biases. *Biology and Philosophy DOI: 10.1007/s10539-015-9512-0*

Potochnik, A. (2009). Optimality modeling in a suboptimal world. *Biology & Philosophy*, *24*(2), 183-197.

Rice, C. (2013). Moving beyond causes: Optimality models and scientific explanation. *Noûs.*

Savitsky, K., Keysar, B., Epley, N., Carter, T., & Swanson, A. (2011). The closeness-communication bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology*, 47(1), 269-273.

Saul, J. M. (2002). What is said and psychological reality; Grice's project and relevance theorists' criticisms. Linguistics and Philosophy, 25(3), 347-372.

Schiffer, S. R. (1972). *Meaning.* Clarendon Press

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 99-118.

Soames, S. (2010). *Philosophy of language.* Princeton University Press.

Sperber, D., & Wilson, D. (1995). Relevance: Communication and cognition. Blackwell.

Stalnaker, R. (2002). Common ground. *Linguistics and philosophy*, *25*(5), 701-721.

Thompson, R. J. (2014). Meaning and mindreading. *Mind & Language*, *29*(2), 167-200.

Tomasello, M. (1995). Language is not an instinct. *Cognitive development*,*10*(1), 131-156.

Tomasello, M. (2008). *Origins of human communication*. MIT press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development:

the truth about false belief. *Child development*, *72*(3), 655-684.

Weisberg, M. (2013). *Simulation and similarity: using models to understand the world*. OUP

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128.