Credit: Penn Today

Lila R. Gleitman

ANNUAL REVIEWS

Annual Review of Linguistics

The Impossibility of Language Acquisition (and How They Do It)

Lila R. Gleitman,[1] Mark Y. Liberman,[2] Cynthia A. McLemore,[3] and Barbara H. Partee[4]

[1]Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; email: gleitman@psych.upenn.edu

[2]Department of Linguistics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; email: myl@ling.upenn.edu

[3]University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; email: cam@seas.upenn.edu

[4]Department of Linguistics, University of Massachusetts, Amherst, Massachusetts 01003, USA; email: partee@linguist.umass.edu

Annu. Rev. Linguist. 2019. 5:1–24

The *Annual Review of Linguistics* is online at linguist.annualreviews.org

https://doi.org/10.1146/annurev-linguistics-011718-011640

Keywords

autobiography, cognitive science, language acquisition, psycholinguistics, syntactic bootstrapping, innateness

Abstract

This autobiographical article, which began as an interview, reports some reflections by Lila Gleitman on the development of her thinking and her research—in concert with a host of esteemed collaborators over the years—on issues of language and mind, focusing on how language is acquired. Gleitman entered the field of linguistics as a student of Zellig Harris, and learned firsthand of Noam Chomsky's early work. She chose the psychological perspective, later helping to found the field of cognitive science; and with her husband and long-term collaborator, Henry Gleitman, for over 50 years fostered a continuing research community aimed at answering questions such as: When language input to the child is restricted, what is left to explain language acquisition? The studies reported here find that argument structure encoded in the syntax is key (syntactic bootstrapping) and that children learn word meaning in epiphanies (propose but verify).

1. BACKGROUND

1.1. Graduate School

From the beginning of my graduate school days, I[1] was exposed to the best of the best among linguists: Zellig Harris and his greatest student, Noam Chomsky, who had just moved to MIT when we met. They represent, at least to my understanding, two opposed trends in the study of language and its learner-users. Harris was the pinnacle of structuralists, believed that what you saw (rather: heard) was what you got, and to my mind is the father of "big data" analysis as it appears in the current computational and psychological literature. For him, the sole legitimate approach was relative-distributional analysis, and his theoretical aim, in these terms, was to find mechanical procedures that moved theory from phoneme to morpheme, thence to phrase, thence to sentence, and finally to connected discourse (Harris 1951, 1952). One obstacle to identifying principles of coherent discourse was that sentences with the same information content could have very different syntactic structures. His big step toward a solution was to render all sentences in a normal or "kernel" form, removing optional variations in form and style while preserving content. The relations by which one could get to the kernel forms of sentences were what was first known as "transformations." One might think of this bottom-up analysis as a proposal about how language acquisition works. That's how I understood it but certainly and emphatically not how Harris understood it. He disdained "psychologizing" and meant to describe "language," rather than the activities or acquirers of such a system.

In my first year of graduate school with Harris, I concretized my own idiosyncratic understanding of the Harrisian point of view in a game—perhaps, more correctly, an evangelical demonstration—which I called The Great Verb Game. It was designed to show that one could recover a specific verb, or small cluster, just from an enumeration of the surface structures in which it appeared. I would think up some "mystery verb" and write it down, covering it. The "players" were to guess it from hearing, essentially, subcategorization frames. So, after a little instruction, the player hears/sees the frame N1 V N2 from N3 and is to come up with an example (say, *take*: *John took candy from the baby*). Next the player was offered another such frame, understood to be a paraphrase or entailment of the first, maybe N1 V N3 from N2. Notice that *take* will not work in this second case (*John took the baby from the candy*); it won't preserve meaning or might even be ungrammatical. So the player revises her guess. After a couple more clue frames, such as N2 V from N3, and N1 V N2 and N3 (and if the players are smart undergraduates or other verbal whizzes), they come up with the (or a) verb that survives this winnowing-down procedure (in the present case, *separate*). Turning over my card, I now triumphantly show them "my" verb, hoping to convince them that the meanings are embedded in the structure in some way. In my first year or two as Harris's graduate student, I tried to formalize these ideas in terms of some derivational web ("chains of Harrisian transformations"), but this turned out to be a false step. In later years, however, these ill-formulated but interesting ideas returned to me in the context of studies in word-meaning acquisition, as I'll mention. But in my first years, this bit of verbal legerdemain so amused audiences that I think this is what caused Swarthmore College to give me my first job.

For the most part, though, my views were rapidly changing because of contact with, and reading of, Noam Chomsky (Chomsky 1957, 1959, 1965). His thinking and findings began to suggest to me, and to a large following of younger scholars who began to surround him, that

[1]The first-person voice is Lila Gleitman's. This piece began as a long interview of Gleitman by her three coauthors, which, together, all four authors edited and revised into this article. We have retained some parts of the original to keep Gleitman's original style.

these surface sentences at the core of Harris's thinking were mere artifacts of an underlying innate system that was only indirectly reflected in our everyday speech forms. Even more importantly, I became convinced of that approach, bolstered by the preliminary findings emerging from my (and collaborators') earliest studies of language in infants and young children. From an autobiographical point of view, I was dragged kicking and screaming toward Chomsky's view, which offended my innately empiricist leanings. Anyhow, I'll now try to say something about the work I did with many brilliant younger collaborators and where it has led me over the years.

1.2. What Do They Know and When Do They Know It?

Few, if any, humans behave like the ideal grammarians that these theories—even Harris's—purport to describe. First there is the question of what babies and toddlers "know" about language despite being unable to understand or produce much of the speech to which they are being exposed. The psychological literature, as it looked when I first examined it, described at most what children actually produced, as a function of age: so and so many words at age one, more at age two, and so forth; one-word utterances at first ("Momma") and two at a time later on ("Mommy sock"). In the early sixties, with the advent of Chomsky's first writings on the psychology of language, psychologists began to write little "grammars" of the child's primitive speech (see, e.g., Brown et al. 1968, Rodd & Braine 1971). But what is the status of these descriptions? Is children's speech, thus chronicled and organized, a reasonable proxy for what they know about their language—or about Language?

With Elizabeth Shipley, my brilliant colleague in mathematical psychology, I set out to probe the toddlers further, asking them, in effect, "Is the following sentence grammatical in your infant dialect?" To be clearer, we tried to find out if *Mommy sock* was an output fairly representing their linguistic knowledge at time of test. After all, when linguists try to build grammars, they never simply use actual speech as the data, but rather, or in addition, use direct or indirect judgments of grammaticality, aiming to get at what constitutes "your language," discarding speech errors, and so forth.

Our finding, in brief (Shipley et al. 1969), was that every kid who spoke in two-word sentences like *Throw ball* acted as if the same two-word sentences from the mother to them were abnormal—they didn't obey those commands, but acted as if something bizarre had happened, whereas grammatical sentences like *Throw me the ball*, they simply obeyed.

This raises a second question. It may be [and Henry Gleitman and I showed, at some length (Gleitman & Gleitman 1979)] that you can't get these judgments in any coherent or full way from even the mature speakers who, if anybody, are the targets that the theory of grammar is, or should be, designed to organize and explain. Most people can't even tell you how *black birdhouse*, *blackbird house*, and *house bird-black* differ in their meaning (that is, in the paraphrases they license). A question we raised concerned how a grammar built from judgments of one population—let us say, literary or otherwise book-wise speakers of some standard dialect—can be used as the empirical basis for building a theory of all native users. Well, we did convince ourselves, in the end, that judgment-giving itself, while useful if available, was not the only relevant source of data.

Yet a variety of uses of language, many of them important, are well correlated with the ability to give linguistic judgments—that is, to contemplate language more or less explicitly, even consciously. As we later showed with several other collaborators, including Paul Rozin and Kathy Hirsh-Pasek (Gleitman & Rozin 1977, Hirsh-Pasek et al. 1978), there is a cluster of skills including learning (or inventing!) an alphabetic writing system, making and understanding puns, and the like that are predicted by this "metalinguistic awareness" (Gleitman & Gleitman 1979). Something like these early generalizations is reminiscent of recent studies and approaches now called executive function, which also attempt to describe and explain certain users' access to aspects of their own cognitive functioning.

2. A FIRST LOOK AT THE INPUT TO THE CHILD: NONEFFECTS OF MOTHERESE

Now we move to the story of what I did in language acquisition, with a succession of brilliant colleagues. Many of them were called students at first, but they never were. They were just younger colleagues.

Penn's psychology department hired me in 1971, even though I had never taken a class in psychology. So I became a member of this excellent department, full of really smart people doing smart things on perception and learning and memory, and also being properly skeptical of linguistics, which I mention because there's much to be said for contact with such "outsiders" for keeping one thoughtful. I still have a visual picture in my mind of the great learning psychologist Bob Rescorla smirking politely at my more overwrought convictions about the nature of language and what we know.

After convincing myself that people know more than they can necessarily say, I still was extremely skeptical of Noam's idealization of a homogeneous speech community—sort of—which came out of that previous work that I did. And the claim that language was innate I couldn't even make out; what could that possibly mean? So I was not only an empiricist at heart, but one of my frequent sayings has always been "Empiricism is innate."

And it certainly was with me. I'd say, What are you talking about? If you're in France, you learn French. If you're in England, you learn English. There's a sense in which it's totally obvious that language is learned from a very precise data set—sentences of English or sentences of French. Language learning comes from the outside in. But it's just as obvious that it's also from the inside out, because many of the cats and the dogs in the house are exposed to the same data set, and notoriously none of them learn English.

So if you're going to study language acquisition, a minimum first step is to distinguish between what's coming from the outside and what may be coming from the inside. An obvious way to do that is to look at what input a child receives. It's not a random sample of English sentences. Does it matter what you hear? Well, obviously, but how does it matter what you hear?

But because all my intuitions are wrong [laughter]—you know, I always start out in the wrong place. And Noam was on about how you go over to Europe on a sabbatical and your kids learn the new language and you don't, so there's something about being a child. And I thought, no, there's something about going to kindergarten and hearing kindergarten sentences instead of hearing learned sentences about linguistics.

With my colleagues, starting with Elissa Newport, we began to look at input. People were already saying that mothers speak a special kind of simplified language to their children, and that's what accounts for the learning.

Henry humorously entitled that kind of language "motherese." [Barbara: He invented the term?] He invented the term.[2] And Elissa Newport, for her dissertation work, began to ask if this "teaching language" really existed in the average home and what, if any, were its effects on infant learning (Newport 1975).

Newport, even as a beginning graduate student, had the clearest vision of how to study, and evaluate, the effects of input variation on the acquisition function. In later years she went on to

[2] Kathy Hirsh-Pasek cowrote a very funny paper (Hirsh-Pasek & Treiman 1982), partly inspired by Henry, when she was my graduate student. This was after Newport et al. 1977. She decided to study how people talk to their dogs and whether that has a causal effect on a dog's learning of English. And she did a very careful study, but she did ask Henry, since he had invented the term "motherese," what she should call the speech of people to their dogs. And Henry said, "doggerel" [laughter]. The field, when it was young, had a better sense of humor, and they actually published this in the *Journal of Child Learning*. But that's not all she did; she does brilliant work in language acquisition.

study the effects of these external influences in unusual and, therefore, revealing circumstances, including second language learning, late learning of a first language (as in some isolated deaf populations), and then how stripped-down artificial languages are used to reveal certain universal or particular properties. More recently she's been responsible for a renaissance of thinking about brain damage and the "critical period" for language learning. Not incidentally, she became my lifelong friend. But back to her first, graduate-school, studies of input and learning—that is, the "inside-out" effects.

In the first study, Elissa simply went to people's houses and talked with the mother, and the little kid was on the floor playing with toys or whatever. Occasionally the mother would say something to the child.

We examined those little corpora with speech to the other adult in the room, Elissa Newport, versus speech to the little kid, to see if they were systematically different—which of course they were. But we also looked at how differences in what the mothers said to their children affected the children's language development. Within the normal range of mothers, at least, few of the variations among the mothers had any effect on what was happening in the children's group—as we measured it. And we measured it in several ways.

In fact, Elissa studied these data to a fare-thee-well for her dissertation (Newport 1975). We were convinced that the maternal style was playing a big causal role in acquisition, so you can't imagine all the analyses that we did. Poor Elissa—she wanted to get her dissertation, but she kept getting null effects. It took us a long time to realize that that was because the facts were not as we all had supposed. That was the first hint about the relative indifference of the learner to the details of the input (Newport et al. 1977).

Today, of course, we would say that the child is so constructed as to build a grammar of an antecedently well-defined kind, no matter what you do. This was a very primitive first step, which at least shook us a little bit loose. And that's all it could do, because it was just an observational, ultimately correlational study. Food for thought, but that's as far as we could get.

3. ACQUISITION IN THE FACE OF INPUT DEPRIVATION

What we decided to do with a succession of further students was to look at cases of much more radical deprivation of input. Assuming that what you learn is an infinite set of relations between forms and meanings, the input has to be forms and meanings in some way. The first thing you would think of is you hear your mother say "Blah blah blah," and a dog walks by. So you learn that *dog* means 'dog.' I have only begged about four hundred questions in so saying, but that's the start of the idea.

3.1. Deaf Children

Now, what if you have a kid who is deprived of information about the forms or the meanings? How could you study that? Well, somebody came from the Pennsylvania School for the Deaf to talk to our seminar about deaf children, and we found out from them—and there's a descriptive literature on this topic—that deaf children who are in homes where there is no sign language begin to make iconic gestures spontaneously. People have studied such gestures, and maybe sign language comes from such original gestures. So this might be a route to look at the origins of language in the child and, maybe, somewhat more extravagantly than I believe this, the origins of language.

Susan Goldin-Meadow, Heidi Feldman, and I had already been convinced by emerging work from [Ursula] Bellugi and others (Klima & Bellugi 1979) that indeed sign language really was a language. We decided to look at deaf children at the stage before they get sent to "deaf school."

In those cases we were looking at, the parish priest would counsel the parents, "Don't gesture to your little deaf child, because if you do they will become too lazy to learn a real language when we send them to deaf school at the age of about five." So here are these little children in your house exposed to no language whatsoever.

I should point out that, as with Newport, Goldin-Meadow's first looks here at isolate language, and deaf language, transmuted into her life's work on gesture and sign; and that she too became one of this growing group of people who formed our circle of thought and continue collaborating to this day.

The first idea was to go to these children's houses and see what we could see, and it's hard to believe until you see it. In one of our videotapes, Heidi shows a picture of a snow shovel. And this three-year-old points to the snow shovel and produces a gestural monologue of how, when it snows, he puts on his boots, he goes outside, he shovels the snow, and so on. And he expresses himself marvelously well.

So there's step one. And you might say, if you were a minimalist, that there you go, we're done, that's language. Or you might say, well, maybe we can do more with these data.

I already believed that every verb comes with its own mini-grammar, as per The Great Verb Game. And if these children had the same logical ideas that you and I have, then they would believe that *give* involves three arguments, and you map it onto language with an NP for each argument, roughly speaking. So *give* and *sleep* and *eat* ought to behave in different ways structurally. Now we asked (Feldman et al. 1978), for the seven deaf kids that we looked at, do they make this distinction?

Now you have to become a psychologist, because life is very tricky—little kids this age produce first only one gesture at a time, then two, then three, then many. But the two-word period of language production is when we had most of our data—*Mommy sock* and all that. So how are you going to find out whether *give* and *eat* and *sleep* have different structures?

Well, let's pretend for the moment—this turns out to be a good pretense—that as far as the output system is concerned, there's this period in which you only have two spaces. You can only say two things. You've got to say the verb. Now, it should be that the required argument always appears as an overt gesture. So where you have *sleep*, everything is fine, but where you have *eat*, each of its arguments will be produced only half of the time, because one space is taken by the verb. You can just do the arithmetic. And for *give*, the proportion should be even lower.

But another thing about language use is that when an argument goes unmentioned it's usually the subject. In fact, this differential omission of the subject NP could easily be seen in the patterns of the deaf children's signing. They "dropped" the subject proportionally more often as the number of required NPs for some predicate increased. Perhaps this is what was later known as "pro drop" in linguistic circles.

That's what we found—for every kid and every predicate. These children exhibit understanding of argument structure, but filtered through this output constraint. Subjects are treated differently from other arguments and words appear in different structures depending on what they mean. And where did that come from? Not from their mothers.

The evidence against my empiricist leanings was starting to pile up here. I was becoming more open-minded. But I always had to fight against my innate empiricism [laughter].

3.2. Blind Children

A couple of years later, along came another brilliant person, whose name was Barbara Landau. She wanted to work on language of the blind. I said to her, "Why would you do that?" And she said, "Well, like you looked at the deaf?" And I said, "Yeah, and later we'll do people with a broken arm?"

And she said, "But there's a reason." Because to some extent they're cut off from the meanings, or at least from the referents. When you say "It's a cloudy day," they're cut off from the clouds.

There's some difference or diminution in your ability to relate the words that you hear to what's going on in the world. And after all, you "must" have learned your language that way—of course, a topic that both Locke (1690) and Hume (1738–1740) had discussed centuries earlier.

So this poor blind child who can't see should be slower in learning and should learn weird things. He should fail to learn some things; we didn't quite know what.

You can predict my first empiricist thought: The child could only learn words for things you could reach out and touch, because touching things has to be their way of getting information. This is what every empiricist should say: You're only going to learn words whose meaning you can get evidence for from the outside world, even though you're blind. So imagine our shock. This changed my whole life. Here's where I fell over the edge. The first verb in this kid's vocabulary was: *see*! With *look* not far behind.

And just as advisors to the parents of deaf children had said, "Don't gesture to your child," clinicians tell the parents of blind children, "Don't use words like *look* and *see* to your blind child, because that will lead to empty verbalism." It's a technical word, *verbalism*. Why? Because, they say, *look* and *see* could have no meaning to the blind.

In fact, when Landau asked the blind child to "look up!", the child raised her hands as if to explore, keeping her eyes and head immobile. This is in contrast even to blindfolded sighted three-year-olds, who in response to the same command will raise their covered eyes skyward. This was a first demonstration (followed by corroborating evidence) that knowledge of sight-related concepts is abstract rather than limited to the visual apparatus in particular, and that a blind three-year-old has means to discover this (Landau & Gleitman 1985). How could they have learned this? Our answer, as it evolved, turned out to be: They play The Great Verb Game, reverse-engineering the meaning from the form. That is, it is the appearance of *look* and *see* with sentence complements (*Let's see if there's cheese in the refrigerator*) that informs the learner that *look* and *see* are terms of perception, not mere contact (one can't say or understand *Let's touch if there's cheese in the refrigerator*).

3.3. Nicaraguan Sign Language: No Linguistic Input

Linguists, including particularly Anne Senghas (Senghas et al. 1997), had the opportunity (and, I might add, the wit) to look at a remarkable case of a language rising in a hithertofore-isolated deaf community. These were children in the area of Managua, Nicaragua, brought together in a kind of informal deaf club. The result, as Senghas and her colleagues showed, was a constantly self-enriching gestural system that in semantics and syntax embodied the elements of known natural languages, spoken and gesture.

I had the opportunity to work with Senghas, Goldin-Meadow, and [Molly] Flaherty on studies of my favorite abstraction, symmetry (and its entailed cousin, reciprocity), in this population (L.R. Gleitman, S. Goldin-Meadow, A. Senghas & M. Flaherty, manuscript in preparation). This work fed several of my obsessions at the same time. It was a new testbed in which an experience-deprived user population showed how abstractions like symmetry and reciprocity arise under untutored conditions and how this delicate formal distinction is reflected in the emerging syntax of their language.

Even the very first user-inventors of this sign language formally distinguished between symmetrical predication (*Juan and Carlos hi-five*) and reciprocal predications (*Juan and Carlos punch each other*). Just like English speakers, these untutored children treat the symmetric predicate *hi-five* as an intransitive verb, with the conjoined *Juan and Carlos* viewed as a single collective subject

argument whose parts are symmetrically involved in the action. In the Nicaraguan Sign Language equivalent for reciprocal *Juan and Carlos punch each other*, one finds JUAN CARLOS PUNCH–GET PUNCHED CARLOS JUAN PUNCH–GET PUNCHED, where the serial verb construction PUNCH–GET PUNCHED encodes a transitive action with two animate participants, agent and patient. The reciprocal construction puts two such clauses together, with participant order reversed, describing the two reciprocal events.

Remember, there was no input that accounts for this knowledge. I take this and the many related findings from Nicaraguan Sign Language to be the most compelling evidence for the robustness of the human language learning function to differences in external experience.

4. SYNTACTIC BOOTSTRAPPING: VERBS OF A FEATHER FLOCK TOGETHER

Coming from the work on the blind child, I had another interesting experience. I said to Henry Gleitman, one of the most thoughtful people I know, "Look! So much for your empiricist ideas! A blind child understands the meaning of the words *look* and *see*." He says, "Well, that's really fantastic. How did she learn them?"

I said, "Heh. You and your—. I don't have to answer that question, really, it's—." And then I thought, Uh-oh. I decided to make a pilgrimage to Cambridge to visit Noam Chomsky, a person of an entirely different persuasion, and a friend of mine, of course, by that time. And maybe I'll get some help in thinking about this. I told Chomsky, "The first verbs in a blind child's vocabulary include *look* and *see*!" He says, "Wow! That's fantastic! How did she learn them?"

Oh my God. When Henry Gleitman and Noam Chomsky are asking me the same question I'm really, really in trouble.

Noam says to me, "Well, it couldn't have been by magic!"

There are a number of questions that really have to be asked at this point. Where did the kid come up with the concept of looking? How do you know that the word *look* is the word for looking—which is approximately 'apprehend' despite the fact that its usage, for the blind, requires touching? What led the blind three-year-old to distinguish looking and touching?

That's when my memories as a linguist, and conversations with Barbara Landau, brought back "chains of transformations" and The Great Verb Game. If you can't learn the meaning of *see* by seeing, perhaps its sense reveals itself in the syntactic structures it licenses. This hypothesis turns on the following idea. If sentence structures are projected from their semantics, then to some extent the semantics itself may be recoverable from the observed surface syntactic forms. Just Harris's position! In fact, such a theory may not be merely a fallback used by the sensorily deprived, but a general clue to the acquisition of word meanings, for almost every word is abstract and mind-driven and requires more than observation of the world to understand.

Indeed, we discovered that the verbs of cognition and perception as used by the mother to her blind child crucially included sentence complements such as *Let's see if there's cheese in the refrigerator*. In contrast, *Let's touch if there's cheese in the refrigerator* never occurred. And it's by using this information that the congenitally blind child learns to "see."

That's basically the answer. You could now reverse-engineer it. That is, if you knew about how languages must map from semantics to syntax, and somebody gave you the syntax, you ought to be able to reconstruct something about the semantics, down to some level.

And that's what we called "syntactic bootstrapping," a term which I invented to mock my good friend Steve Pinker's "semantic bootstrapping" (Pinker 1984). Pinker, who was right in many ways, was saying that you can acquire the syntactic structures by understanding the semantics.

But everything I was doing was saying you can't, because that semantics business that he thinks you should have first isn't so easy. While Pinker's premise was false in part, there's something deep in his work about how the syntax–semantics correspondence can help you learn semantics. We get the main clues from what we learned from the deaf and the blind—that you can exploit that correspondence from the other end. At the point where verb learning is going on, enough syntax is accessible for the child to perceive. And from that, if you're a creature endowed in a certain way by nature, you can make a very good guess about the semantics of the verb in a sentence.

Here is an important detail. A verb will occur in several syntactic environments, some with varying numbers of arguments. So one can say, *I gave a book to John*, or *I gave at the office*, or *I gave cash*. But the maximal number of arguments semantically partitions the verb set in useful ways.

Of course this is complicated. But you know, infants aren't just infants; they're smart [laughter]. Right?

So the relationship between number of arguments and number of NPs—there's something natural about it. And the same is true for the relation between sentence complementation and words like *look* and *see*, since we can perceive both objects and states of affairs. Therefore, just as in the original Great Verb Game, listener-learners require several framing structures to retrieve the semantics of individual words.

[Mark: Is it fair to say that your syntactic bootstrapping theory is sort of a redemption of Zellig Harris's syntactic discovery procedures on the basis of transformational relations among sentence frames?] Right! Exactly. Because as a child, you don't get to hear underlying structures. You hear some form of a surface sentence. And you have to do something with it. So, yes, I think it's basically Zellig Harris rediscovered, redeemed, and put to purposes which he would hate [laughter].

But how else would you really understand Harris's theory? Here's how Harris expressed it: "You know, you really put together a grammar by intuition. But to have a theory you have to show that it could be mechanically invented." You have to show that you could have done it algorithmically. A machine could have invented it. That was his story. And that machine which could have invented it is the theory of language.

Well, I believe that too, but I believe that's the language acquisition device, fondly referred to as Universal Grammar.

5. TAKING THE THEORY INTO THE LABORATORY

Let me tell you how we now bring this into the laboratory to make good on all of these claims. We see this apparent relationship between argument structure and verb meaning, and we want to know, first, whether these patterns hold across the lexicon and, second, whether big people and little people implicitly know and use these patterns in the surface syntax to understand what the verb means. Can you play The Great Verb Game across the lexicon, not just with your favorite examples—and do language learners play The Great Verb Game as they're learning?

5.1. Structural Cues to Verb Meaning: Syntactic Clustering and Semantic Clustering

The first work was by Cynthia Fisher, who was lead author on a paper with Henry and me in 1991 (Fisher et al. 1991), which demonstrated that verbs of a feather flock together; that is, verbs whose meanings are judged to be closely related appear grammatically in highly overlapping sets of structures. This work was the beginning of Fisher's illustrious career in language acquisition, in many ways explicating and concretizing the ways that syntactic bootstrapping might actually work.

Figure 1

A sketch of a scene shown on videotape. The actual characters were actors dressed in rabbit and duck costumes. Drawings courtesy of Henry Gleitman.

5.2. Verb Acquisition and Frame Compliance

As this general line of research was beginning, we did an experiment with Letty Naigles (Naigles 1990, Naigles et al. 1993). You show two- and three-year-olds a Noah's ark, and on a little stage they have to act out *Noah brings the elephant to the ark*. But they also have to act out *Noah comes the elephant to the ark*, where now you've added an extra argument to *come*.

What do the children do? They could either think you're crazy, or they could act as if *come* means 'bring.' And indeed the children treat *come* as 'bring,' a transfer verb, when you add an argument. Similarly, they treat *bring* as 'come' if you subtract an argument. This happens very systematically. We call this frame compliance—compliance of your interpretation with (the semantic information latent in) the syntactic frame. And Letty Naigles had a study of even younger children that also showed a robust syntactic bootstrapping effect (Naigles 1990). You have babies of age approximately two watching two video screens, with both screens initially showing the same video. **Figure 1** shows a sketch of an example. We filmed people dressed up in costumes, here a rabbit and a duck. What you see is the rabbit pushing the duck down into a squatting position, while at the same time the rabbit and the duck are wheeling one arm in a great circle. So there are two events happening in the world that the children observe.

As the children watch these things, half of them hear, "Oh look, the bunny is gorping the duck." And the other half hear, "Oh look, the bunny and the duck are gorping" [laughter].

Now the screens change (**Figure 2**), and on one screen you see the rabbit still pushing the duck down, but no arm wheeling, and on the other screen they're standing next to each other, still wheeling an arm, but the rabbit is not pushing the duck down. "Oh! Now this is different! Find gorping now." And you observe which of two screens the children are attending to.

Almost every child shows the effect. The ones who heard the transitive sentence now looked more at the causal pushing scene, and the ones who heard intransitive training sentences looked more at the arm-wheeling scene.

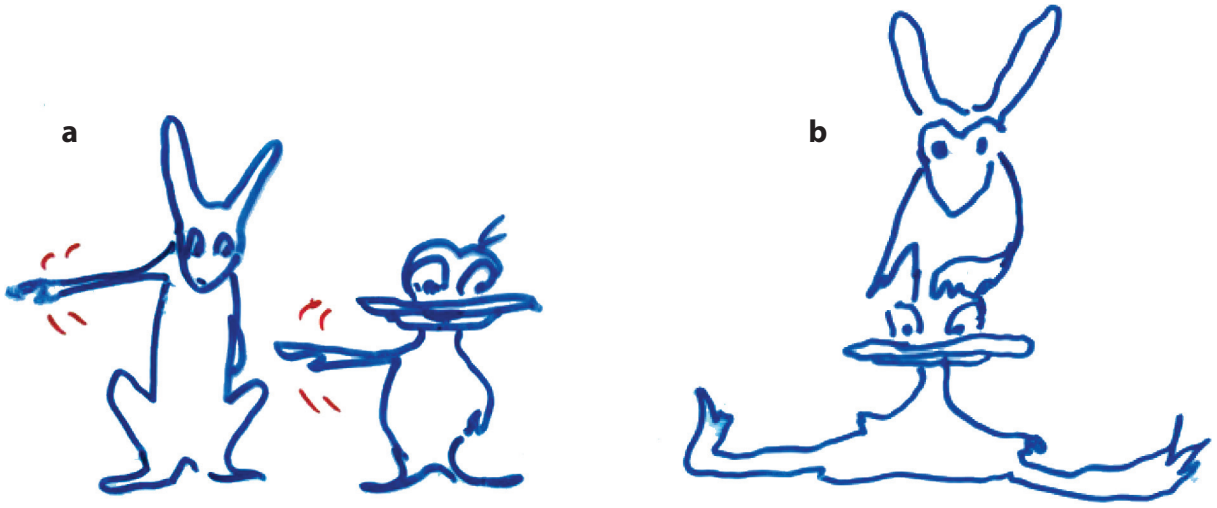

Figure 2

A sketch of two videotaped scenes shown on separate screens. The actual characters were actors dressed in rabbit and duck costumes. Drawings courtesy of Henry Gleitman.

So in the presence of two events, the syntax narrows down your attention. That's all the syntax can do—semantics is much richer than syntax. But there's enough information in the syntax to point to the right neighborhood for the meaning of the verb. And then "the world" has some hope of supplying the detail.

5.3. Not All Cues Are Created Equal: Syntax Trumps Morphology

In English if you look at the number of NPs it's quite a reliable cue to number of arguments (as we just discussed in connection with Naigles's work). But what if you look at languages which allow extensive argument dropping at the surface, so that that cue is less reliable? And what about languages that make greater use of morphology to mark argument structure?

Jeff Lidz, a very talented and creative linguist, became a postdoc at Penn. He studied Kannada, a language that differs typologically from English in its use of these cues to argument structure. Kannada has a causative morpheme (compare English *-ify* as in *purify*) that appears whenever a verb with causative meaning is expressed. In Kannada this morphology is more reliable than number of expressed arguments as a cue to causal interpretation. Would a child just use whichever cue was best in his/her own language? Or was there some tacit priority, not all cues created equal in the child's mind?

As Jeff showed, Kannada-speaking two-year-olds in the Noah's ark procedure nevertheless rely very heavily on the number of overt NPs and ignore the reliable causative morpheme in their own language (Lidz et al. 2003). That became the first crosslinguistic playing of The Great Verb Game.

We had the good fortune of having Jeff here for a couple of years because he became a hybrid character—the linguists came to say he was a psychologist, and the psychologists that he was a linguist, which as he pointed out, made it extremely hard to get a job. Fortunately for us, he was stuck here until his kind of work and thinking became established as cognitive science.

5.4. The Acquisition of Belief Verbs: False Belief and Syntax

There was an experiment with mental verbs done by Anna Papafragou, Kimberly Cassidy, and me when Anna was a postdoc at the Institute for Research in Cognitive Science. Anna, too, has become a cognitive scientist renowned for studies of acquisition and the interaction of semantics and pragmatics—and also, of course, a warm friend and continuing colleague. Kimberly tops it off by becoming the president of Bryn Mawr College.

The name of our paper was "When we think about thinking" (Papafragou et al. 2007), because here's a funny thing: It's easy to teach adults or three-year-olds that *gorp* means 'think,' as long as you put them in a dark closet where there's no world. As soon as there's a world, we're in trouble. Even for adults, if you show them Rodin's statue and you ask what he's doing, or what's happening, or you try to ask this in the most neutral way, they say, "Cupping his head in his hand." And you say, "Idiot! The name of the thing is 'Thinking'! It's the Thinker" [laughter].

But people don't naturally observe thinking as thinking. So how does anyone learn the meaning of the word *think*? We can show in the laboratory that if you just say *gorping* in the right syntactic environment, they'll guess 'thinking,' but it seemed very unlikely to be what actually happens.

Kimberly Cassidy, who was also a member of our seminar, had a funny incident with her little son. The family went in to get a pizza while she stayed in the car, and the kid comes out holding a cardboard cup that says *Coke*, *Coke*, *Coke* all over it. The mother says, "Put that down. You know you're not allowed to drink soda." The kid looks at her and he says, "Oh, this is water; the man gave me some water to drink." The mother says, "I'm so sorry. I thought from the cup it was Coke." And the kid says, "Oh, that's so funny. You were mad. Let's do more."

We had talked about this a lot, because Kimberly used the Human Simulation in her classes at Bryn Mawr as a demonstration experiment for her students. And she said, "I got it! Now I know when 'think' rises into your consciousness and you think about thinking. It's when you're in a situation of false belief or trickery."

So we designed an experiment which contained situations of false belief and true belief. They go something like this. A woman is sitting at the table drinking her tea and reading the paper with a teapot nearby. Another woman comes in and wipes off the table, moving the teacup and teapot. In the true-belief situation, the woman picks up her teacup and drinks, but in the false-belief situation she picks up the teapot and almost drinks from it. We asked people of 3 years of age or of 23 years of age what happened.

In the situation of true belief, they just say she was drinking her tea. In the situation of false belief, they often say, "Oh, she thought that was her cup." So the situation of false belief causes you to think about thinking and talk about thinking.

So we crossed that with the syntax. We showed the kids a video. A little boy is walking to his sick grandmother's house with a basket of food for her. In the true-belief situation, a cat offers to help him deliver the basket, and does so. In the false-belief situation, the cat runs ahead to hide the grandmother in the closet, then gets in bed pretending to be grandmother so as to get the food. At the end of the video, the cat says a sentence with a nonsense verb, and the experimenter asks the child to help her understand the cat—kids love to help as interpreters. The cat's sentence may use either a transitive frame ("Did you see that? Matt *gorps* a basket of food!") or a sentence-complement frame ("Did you see that? Matt *gorps* that his grandmother is under the covers!"); this is our other variable.

Now, think of that as a two-by-two—two situations, two syntactic frames. In a situation of true belief and a transitive verb, you get a belief verb as the meaning of *gorp* approximately 0% of the time from adults or kids. If it's a false-belief situation, you raise the probability of a belief verb to about 12% for kids and under 10% for adults. It's not huge, but the false-belief situation does raise the salience of 'believing.'

And if you put the false-belief situation together with sentence-complement syntax, the adult subjects are practically at ceiling, with 80–90% belief verbs. Kids guess belief verbs 40–50% of the time—not as much as adults, but far more than when they get just syntax or just the scene. So, again, it's a story about convergent cues. The world—the false-belief situation—is contributing something. But either database is insufficient alone. And that brings us to *chase* and *flee*.

5.5. Bias, Hard Words, and Even More Evidence for Syntactic Bootstrapping

You can't tell *chase* from *flee* by looking at a scene. And you can't totally tell *chase* from *flee* by looking at the syntax, because the subject could come after the object in some languages. But from the syntax and the scene taken together, it's determined. That's what our many *chase/flee* studies show (Gleitman et al. 2007, Nappa et al. 2009).

People in psychology are always talking about relevance and eye gaze, and I don't want to denigrate those things; they do have an effect. If you give people a *chase/flee* picture in an uninformative syntactic environment, there's great bias toward *chase* over *flee*; let's say 80% of the time they say the rabbit is chasing the elephant. You can strengthen that bias by having them looking at the chaser, to 90% or 95%. But if you give contrary syntax—if you say the elephant is gorping the rabbit—there's no effect of the eye gaze whatsoever, even in three-year-olds. The interpretation is *flee* 'run away from'—the syntax completely overwhelms the situational cues.

6. THE HUMAN SIMULATION PARADIGM

Now comes what I call the Human Simulation Paradigm: using adults to "simulate" child learners, first described by Gillette et al. (1999). These experiments introduced a novel methodology that allowed us to see what kinds of information are contained where in the input, working alternately with "human child-simulators" and children.

Anne Lederer married and became Annie Duke and, after doing the research I'll describe, left to win the World Series of Poker and to write and lecture worldwide on topics of risk and reward in poker and life. Still my close friend, she is plotting to finish her PhD at Penn, after which Henry can be heard saying (he always did, but now it can be more literal), "Doc Holliday, famous card-player but a Penn-educated doctor too."

We called it the Human Simulation Paradigm in analogy with computer simulations. If people complained, "You're looking at adults, not children," we could reply that adults are more like children than your Sun Microsystem is. If you believe in machine models, you ought to believe in my using human adults as the subjects. We set up conditions for adults—who were easy and inexpensive to work with—that in various ways simulate the learning situation that children are in. When we get interesting results, then we try it with children.

6.1. Background

Here's a story about language acquisition. First, here you are, a baby, and you don't know any syntax of English yet. You can't start syntactic bootstrapping if you don't even know where the subject of the sentence is. The first thing that you can do is, you hear *dog*, and you look wildly around the world to see if you can find some behavioral concomitant, something out there that occurs when the word *dog* occurs.

There's a lot of good information around now that indeed there is such a stage. That's what I call "word-to-world pairing." It's terrible. It's not very good (see Section 7), but it does a few things. What you can learn from that kind of information is a few whole object term nouns, and

very little else. But, crucially, you can learn words that are used just when there's a co-occurring referent.

Suppose the story is that, first, you learn a few of these nouns. They're going to give you an anchor; they're at least going to tell you where the subject is in your language, so you can make something of the rudimentary surface syntax that you're hearing. It also makes syntactic bootstrapping available. That is, this first learning gives you clues to the rudiments of the surface structure, to the extent that this varies crosslinguistically—your passport, in a way, from "I know they're speaking some language" to "It must be English."

Now you can do structure-to-world relations in addition to, or instead of, word-to-world relations. Why should this help? For reasons I've been discussing throughout: The structure provides information about the semantics of the system. This move, from understanding words to understanding words in their sentential contexts, sketches an incremental language development theory.

6.2. The Experiment

Now I'm going to turn the adults into learning children in various ways, on the following hypothesis. For a limited set of words, most of the information is in the world for you to see, like the difference between bouncing and rolling. If the information is in the world, God says, "Go out into the world, my child, and learn the meanings of words by observing the real-world contexts in which they're used. But if you have any trouble, come back. I'll tell you another method. And that second method will be: Look at the linguistic licensing of those words, their syntactic frames and selectional co-occurrence. And that will tell you the answer."

Here's what we did. In this procedure, called the Human Simulation Paradigm, I'm going to give the subjects, Penn sophomores, information of one of three kinds and combinations thereof. All the stimulus materials derive from actual videotape of mothers at play with their 14- to 18-month-old language-learning offspring. The three kinds of information subjects get, alone or in combination, are (*a*) "world," or "cross-situational observation" (i.e., video clips); (*b*) noun co-occurrence, or "selectional" information; and (*c*) syntactic frame.

6.2.1. Part one: Some nouns can be learned from observation, but verbs not so much. In part one of six parts, the only thing subjects who represent the babies get is the sound of a word, or a nonsense syllable, or a beep, along with "the world"—the context in which this word is used. So, you're going to get a videotape of one minute in length, and it shows you a mother and her 14-month-old baby, and during that minute, chosen at random from a huge corpus of mother–child speech—chosen at random; that's very important—something is happening.

The original soundtrack is muted, but at a certain point during the video you hear a beep, and that's when the mother says the mystery word, which it's your task to guess. Think of that as learning your first word. But maybe you don't learn from one trial. So we show you six different cases, all chosen at random, of mothers using this same mystery word.

In the experiments there are about 12 mystery words, and they're all interleaved. So you don't hear six cases of *hit* followed by six cases of *run*—because mass learning has its own issues. You see a *hit* scene, and you see a *run* scene, and so on; they're all mixed up. And you hear, let's say, a nonsense word with each. You're trying to learn these nonsense words.

Now, we have about 100 undergraduates, and they try to guess it. You can think of this as crowdsourcing the probabilities of one word or another, given the information available.

The first thing you realize from these experiments is why deaf people stop watching television—because it's very hard to figure out what's going on in any detail without the soundtrack. And don't

forget, you should think of this as learning your first words. You don't know any syntax; at least you don't know the surface syntax of the language you're hearing. You're going to have to try to guess a word just from its context. You think that's easy?

[Barbara: This gets rid of an argument I used to imagine that I had against Chomsky. He said, "Syntax is completely autonomous. You don't need any semantics." I said, "No, people don't learn language from listening to the radio."] Well, you were both right. You're going to have to find out what the referents are to some of these words. It will matter that you hear *dog* when there are dogs around. The question is: What kind of information does exist out there in the sheer co-occurrence of a scene with a sound? On Chomsky's side of that dispute, neither TV nor the radio (that is, observation) is going to reveal the meaning of *probably* or *think*.

We see that the answer is complicated, not simple. What can you learn? It turns out that if you choose from about, let's say, the 10 most frequent whole object nouns that little children learn, your subjects get about 40% of them right in this kind of task. But they only get 10% of the most frequent verbs right. No matter how many examples you give them, they not only don't get the verbs right, but their answers don't converge. They're not getting warm, so to speak. In contrast, they're pretty good on these whole object nouns of the very simplest set. So, 40%—that's not great, but it's not terrible. Here's one of the most robust effects in the language acquisition literature, crosslinguistically as well as within English: Children learn nouns before they learn verbs.

So there's this one-word period; it's got practically all nouns. Not all, importantly, but mostly whole object nouns. The input does not begin to account for this disparity. Of course, you do use more nouns than verbs, but the percentages are nowhere near right.

That's part one of the experiment.

6.2.2. Part two: Even given noun co-occurrence information, verbs are still hard. All the remaining parts of the experiment look at how verbs can be learned once you know some nouns.

First we tested a "noun co-occurrence condition." Suppose you've learned some nouns but nothing yet about syntax. Does just hearing two nouns in random order help you guess what verb they occurred with? Well, if you heard *cake* [and] *chef*, then it's more likely to be *bake* than *punch*.

We tested this with paper and pencil, extracting nouns from the same six sentences for each verb that we used in the first part of the experiment, and having subjects guess the verb.

Can you guess the verbs? Well, again, you're down around 10–14% correct—it's very hard—about the same percentage as you were able to get from the video alone.

6.2.3. Part three: syntactic frame and convergent cues. We now return to Lewis Carroll. Or Zellig Harris. Or Lila Gleitman pretending she's Zellig Harris. We suppose that the kids have learned a little bit of syntax, and we give them the syntactic frames. It's like Jabberwocky, normal syntax and morphology, but nonsense words. We start from the same six sentences we used before for the given verb, but we convert the nouns to nonsense words and the verb to a nonsense word in capital letters: *Can ver GORP litch on the fulgar?* (The original sentence was, *Can you call Markie on the phone?*) And we say, "Guess what word this was that Mommy is saying? Guess the verb." In the previous conditions, they had six visual scenes, or six pairs of co-occurring nouns; now they have six sentence structures, six syntactic frames, and nothing else.

Here the results are very different. First of all, subjects are very good at this in a funny way. For the abstract verbs of belief and desire and perception, they're about 50% correct. For the concrete verbs like *jumping*, they're worse: about 10 or 12%.

[Barbara: Is it because there are too many verbs that have the same frames among the concrete ones and not as many mental verbs?] Yes, that's right. In part it's that there is a comparatively small

number of mental verbs. But what also contributes is that sentence frames that imply something mental are more distinctive—that puts you in the neighborhood.

And I'll tell you, this experiment didn't make the differentiation you want very well, but we had that other great experiment on the acquisition of mental verbs by Papafragou et al. (2007) that we discussed earlier. This goes back to my story about the efficient rather than the good god. A good god would give you all the information to learn all the words, right? In the same way. But instead you have an efficient god, who says, "I'll give you a certain kind of information, and that's good for learning a certain kind of word. But for the other kind of word, abstract verbs, I won't give you very good information about the world. You can't tell thinkers from looking at them. But I'll give you this other kind of information instead." It's not nonoverlapping, of course.

But the way that you change how much work the syntax is doing in Jabberwocky is a function of the abstractness of the verbs. I mean, they're not wildly abstract. We're talking about words that get used by two- and three- and four-year-olds. So it doesn't include *suppose*, although it does include *think*.

6.2.4. The rest of the experiment, and how I became a Platonist. Now you can start to put these three conditions together—the scenes, the co-occurrence, and the syntactic frames—and we had more parts of the experiment that did that. If I give you the syntax plus co-occurrence, I'm giving you the whole sentence except for the mystery verb: *Can you GORP Markie on the phone?* You get very high results then—the effects are definitely overadditive. With different combinations, you get somewhat different answers.

The more concrete the verb—that is, the earlier it's learned by little children—the more of that verb is given in the world and the less of it is given in the syntax. So we say, no, *think* was never too abstract for a baby. Babies can think about thinking. If they couldn't, how would they learn to?

I'm a Platonist by now. I've gone from one extreme to the other because of these children. They knew it all in advance. But as Plato [2009 (ca. 385 BCE)] points out—there's this great line in the *Phaedo*—he says (75e), "But if the knowledge which we acquired before birth was lost by us at birth, and afterwards by the use of the senses we recovered that which we previously knew, will not that which we call learning be a process of recovering our knowledge, and may not this be rightly termed recollection?"

I think the important thing for me, at least, is I didn't start out believing crazy things like this. I kept saying no, and the world kept contradicting me—or, I should say, the kids kept saying, "Yes I can. Yeth I can" [laughter].

But maybe I'm a little bit different from Noam in this: I'm extremely impressed with the capacity of human beings to learn. I think that's how we got in all this trouble with empiricism. Because you'd have to be crazy to think that humans aren't richly endowed differently from all other animals with the capacity to learn even the craziest things. And that is such a remarkable and outstanding property of humans that you tend to de-emphasize how much is not learned in that sense.

7. EASY WORDS: FIRST STEPS IN ACQUIRING THE LEXICON

I had the opportunity, starting about 20 years ago, to begin to talk with and collaborate with a wonderful young psychologist named John Trueswell. We started off with quite different preoccupations. He studies processing of language, and I studied acquisition, but the more we worked together the more we saw these as simply different aspects of the same overall story.

Figure 3

A plausible word learning environment for the word *shoe*. Adapted with permission from Medina et al. (2011).

7.1. Background: Aggregate (or "Cross-Situational") Learning

Trueswell and I, with several collaborators, began to look in earnest at the initial word-to-world procedure. It's clear that any observation of the scene in view makes available multiple conjectures about the meaning of a word then said, even if we presuppose that it will be something relevant. (Try learning the meaning of *shoe* in the ordinary environment shown in **Figure 3**.) In light of this indeterminacy, it's commonly thought that multiple observations can redress the problem. So, one would hope—although one sees different dogs in different times and places doing different things while hearing the word *dog*—that the utterance of "dog" should appear more often in the presence of dogs than in their absence. Aggregating these instances could isolate the dog as the intended referent when one hears the sound "dog."

Oversimplified as it is, that's the widely accepted story of incremental associative learning. There is a problem here, however. To preserve in some relatively unanalyzed way the complexity of observed scenes is already a heavy burden on any memorial scheme one can think of. Comparing across any two or three or four of these scenes, then, might create a computational explosion rather than the desired convergence. Not only are there problems in principle, but there is the wee problem of describing actual child word learning where we see, again and again, children learning from one or a very few encounters the meaning of a new word, what Susan Carey called "fast mapping" (Carey 1982).

In recent years, very interesting work from Linda Smith and her collaborators (Yu & Smith 2007) seems to support the aggregate or gradual learning of word meaning even so. In their work they presented listeners with successive exposures to a word and showed that the group performance improves as trial number increases. But there are two ways to think about such results, as has been recognized by past commentators as well. Let me tell you about it using a mock example from Henry Gleitman, which I think does it best.

As Henry put it, consider a lightbulb. In the good old days of lightbulbs, you buy a lightbulb and at some point it goes out. It doesn't get dimmer and dimmer. It doesn't gradually learn to go out. It learns on a single try to go out. That is to say, it goes out someday.

When does it go out? Well, you never can tell exactly, because it depends on things about the filament that you can't see. If you look at a group of lightbulbs over a period of time, you see an ascending likelihood of failure; the lightbulbs as a group gradually go out. But it's not true of any single lightbulb that it gradually goes out. It's a one trial go-er out-er.

This whole process has been distorted by pooling data. What the group does may not be characteristic of what any individual does. Every subject may be a one trial learner. But if you look at the group over time, it's going to look like you have gradual learning.

The aggregative position has been dominant in psychology, and here it was extended to word learning. But you have to look at the interim stages for individual subjects. What was the trajectory of their learning? The short story is that subjects act very much like lightbulbs, acquiring a term more as a one trial epiphany.

The evidence looks like Hume's intuition about primitive learning: the association of entities that occur contiguously in time and space, with a little cause and effect mixed into the bargain. So, you show them a duck under the right circumstances, they learn *duck*, and that's the end of the story: They know *duck* for the rest of their lives. They don't need a whole duck pond full; they don't need big duck data; they just need the right data. Of course, this raises the question of what makes data right. What makes data a learning opportunity?

7.2. Reviewing the Human Simulation Paradigm Results for Evidence

The first clue Trueswell and I had was when we went back to the findings of the Human Simulation Paradigm, the part where they just see a one-minute video, and they just hear a nonsense syllable or a beep and they have to guess the mystery word. They do variously well depending on what kind of word it is, and at best, if you take the simplest whole object nouns they can guess them 40% of the time. What's going on there, that they can sometimes successfully guess it?

What we found is that there are some good observational circumstances. For most stimuli, the subject is guessing correctly 3% of the time—2%, 4%—and then there's a stimulus where 50% of your subjects suddenly guess it correctly. And they more or less hold onto that. But it's like there was a circumstance that apparently gave them hope of learning. So, there are good stimuli and bad stimuli.

Maybe what's going on is, yes, we gave them six opportunities, but they really only learned from one—the good case. There's room to at least investigate that hypothesis.

Here's what we think. It would be impossible to learn if your thought was that people store the situation with the word. But here's what you could do instead. If you see a plausible situation for word learning, you make a guess about the word meaning, and you store the guess. You throw away the situation.

So, you say, *mipen* means 'horse.' You have a proposed lexical entry. And now you wait for another observation. When you hear *mipen* again, if there's a horse there the second time, you say, Okay, that's it: *Mipen* means 'horse.' The determinative, even absolutistic way I've described this might make Trueswell a bit uncomfortable; he would not disagree with this rendition, but he sees it in more realistic, probabilistic, terms.

We call this story "propose but verify." It's basically one trial learning. Or, as I think of it, one and a bit trial learning.

7.3. Human Simulation Paradigm Experiments Testing "Propose but Verify"

Our group designed some experiments (Cartmill et al. 2013, Medina et al. 2011). We had some good fortune and got new data.

One of our collaborators, Susan Goldin-Meadow, at the University of Chicago, had been, for entirely different purposes, collecting information—going to the houses, talking to the kids—on 63 socioeconomically stratified families in Chicago—ordinary kids, upper class, middle class, and so forth. Her team of experimenters visited once a month, for an hour or so. All the talk was videotaped. It's massive amounts of data. But it was all transcribed, so we could search and select from that database. In our experiments, each input was a real mother's one-minute video clip randomly chosen, of saying the word *shoe*, let us say.

On each of a dozen or so very common words, we got about 10 clips per family. We ran these thousands of Human Simulation Paradigms, and the first thing we found was that there were some good cases where subjects guessed well as a group.

In a certain video of *shoe*, the subjects were good at guessing. But even though we were selecting very easy nouns, only about 1 in 12 observations does the subjects any good. The rest are complete garbage. It's not just that you might not be able to learn *shoe* from them; you might, God forbid, learn something else and have to back off from it, because it might be a good case for learning *socks* or *apple*.

So the situation indeed looks terrible. But nevertheless, about every twelfth one there's a jump and the subjects get it.

To make a long story short, these experiments gave good evidence for one trial learning, with the "propose but verify" strategy, as opposed to the aggregative learning procedure which requires storing the multiple conjectures made, whether right or wrong, and the situations they were made in, and doing some processing on all this stored information. Our finding was that unsuccessful guesses are not stored, nor is the situational context they were made in, because the irrelevant and misleading information is filtered out.

7.4. Laboratory Experiments Testing Adults and Children on "Propose but Verify"

The next thing we have to do is notice all the things that are totally uncontrolled and uncontrollable in the Human Simulation Paradigm where, after all, we used clips from real videotapes. We have the advantage of this being real life, a real situation. It isn't a picture of five objects arranged in a circle. That's good, but on the other hand it's also bad, because you don't know what in those situations is really leading to the effects we found. The aggregative learning researchers (Smith et al. 2011) suggested that some of the differences in experimental methods were leading to the differences between our findings and theirs.

For the work reported in Trueswell et al. (2013), we decided to do "the perfect laboratory experiment," where we could really expose the logic of this thing. We set it up just as in Yu & Smith (2007). Now we have five objects arranged in a circle (**Figure 4**). Let's look at the logic, because it's completely transparent here.

You see a series of pictures each with five objects, and you're to guess each time which one *zud* picks out, with no feedback about whether you got it right. Here the "right" answer is 'bear'; a bear appears every time the word *zud* is heard. So, on your first trial you guess correctly 20% of the time. Now, if you guessed 'bear' the first time, and you see that there's a bear in the next trial, your rate of guessing correctly goes way up. But if you guessed 'door' the first time, what happens when you get the next stimulus, where there's no door?

Remember that on the cross-situational aggregative learning story, accruing and storing knowledge about the *zud* situations, you guess better than 20% on the second trial, because there had been a bear on the first trial too, even though you didn't guess 'bear' that time. But that doesn't happen. The results were that if you guess right on the first trial, you go on and guess righter

Figure 4

Example sequence of two learning trials for the word *zud*, which meant 'bear.' Adapted with permission from Trueswell et al. (2013).

and righter. But if you guess wrong on the first trial? Then the next one is just like any trial in isolation—you do no better than chance.

That is to say, you totally forgot everything except your initial conjecture. So it does you no good that there was a bear there on the previous try. We went on and showed the same results when there were sometimes only two choices in some of the pictures, and very close to the same results when there were only two choices in all of the pictures.

That was nice, and it looks like a good proof of one trial learning. But it was all done using the Human Simulation Paradigm, and there comes a time when you might say that it's not enough to show that this is the behavior of undergraduates at the University of Pennsylvania. Because we know that children are very good at word learning, and language learning, and we have no idea whether this kind of theory really applies to children. The child-friendly versions of Woodard et al. (2016) show the same pattern.

8. CONCEPTS

In this so-long interview and discussion of my work, I never really mentioned—but now will—how restricted the topics are overall. What I have meant throughout by "learning the meaning of the word" is discovering the mapping between a concept and the expression used to refer to it in the child's exposure language.

But that all leaves aside the central question of what a concept or category is in the first place, a topic I have studiously ignored. There are, of course, very good reasons for ignoring it because it has formed perhaps the central question of great minds, starting, most notably, from Plato and on into the present era.

However, I was not always able to totally evade thinking about the problem of conceptual representation, and looked at it in a couple of spotty ways. As you know, fashions change about what a concept is. Some think concepts are mental images. Others think concepts are a set of necessary and sufficient conditions. And more recently, some have thought that concepts are a set of typicality conditions for the use of a word or expression.

In two articles, we looked at these issues from the point of view of compositionality. Both of our contributions had to do with prototype theory, and both were largely negative, as perhaps won't be surprising if you look at the list of authors.

The first article (Armstrong et al. 1983) showed, or purported to show, that word meanings do not decompose into their prototypes. The second (Connolly et al. 2007) showed, or purported to show, that words do not compose (into phrases) over prototype representations. Why is the issue

of compositionality so important here in deciding on what concepts might not be? Suffice in the present discussion to say that it is because we do not speak in one-word sentences, and we do not have to do more language learning to understand phrases never before heard, including ridiculous ones like *purple cow*.

Returning to the topic of what I myself have done and not done, I have studied word and concept learning in the absence of a clear picture of what concepts are in the first place. I don't suppose that pointing to general ignorance on this topic is a good excuse, but there you are. So I have studied only the mapping problem, how you decide that [dog] is the way they say *dog* in English, a very restricted but investigatable set of issues.

9. WHORF: DO CONCEPTS COME FROM LANGUAGE ITSELF?

I have just asserted, very truly, that neither I nor probably anybody else has a very successful theory of what concepts are, although I have had something to say about what they are not. This general topic raised my interest in Benjamin Whorf's theory of the relation between language and thought (Whorf 1956). After all, a good way to think about what I've always tried to understand is where knowledge of language comes from in the first place.

Whorf, at least on modern readings, more or less turns this problem on its head and argues that thought is derivative from language knowledge. How strongly you want to assert that is, of course, a question of which paragraph or sentence of Whorf's you want to cite. But the general notion that concepts arise from our linguistic experience is something I've always thought congenial to the way most psychologists think. The idea that concepts are some inborn, unlearned atoms of thought is an uncomfortable position for any serious empiricist. Yet every issue of developmental psychology reveals yet another abstract notion present and available to 6- or 10-month-old infants. Where is this conceptual knowledge coming from? And how could it appear so dramatically early in life? Well, one source of data that has been deluging infants since conception are the data of the natural language they are hearing. Perhaps this is the real "outside source" for concepts and their organization. This is all wrong, but I have always thought that that must be why psychologists like Whorf's position so much.

My brilliant young colleagues Peggi Li and Anna Papafragou and I began to look at a series of crosslinguistic experiments on spatial organization. In an elegant line of experimentation, Levinson (1996) had tried to show that spatial representation and spatial problem-solving differed crosslinguistically as a function of the lexical resources each language makes available. People were said to solve spatial puzzles differently depending on whether their languages had such egocentric spatial terms as *left* and *right* in addition to the universal allocentric terms of cardinality such as *east*, *west*, and so forth.

What I believe we showed (Li & Gleitman 2002, Li et al. 2011), to the contrary, was that these language-specific effects were products of coexisting cultural differences and very particular manipulations done in these investigators' laboratories. Correcting for these, the universality of spatial cognition itself was reemphasized in our work. Not that either position has been accepted in the relevant communities of psychologists, anthropologists, and cognitive scientists—rather, on this kind of issue everybody has managed to remain firmly where they were before but slightly shifting the supportive experimental work cited in its favor. The issues are just too deep.

10. CONCLUSION

In many ways, I think of my lines of work as (partly inadvertent, partly adventitious) elaborations of several of the approaches to language and mind that Chomsky and [Jerry] Fodor brought into

the study of language, thus creating a more general revolution in how cognitive science thinks about the nature of mind. The most general conclusion pertains to how much and how widely knowledge is a function of the structure of mind and how restricted the influence of experience is. This is why the many studies of acquisition in infants with deprivation of input showed so little effect on the final product.

At the same time, there is much to be said in support of Zellig Harris's perspective, which looks so different on the surface. After all, Harris held, and so did Plato, that the knowledge we had at birth is forgotten. I interpret this to mean that although we are surely endowed with a language capacity, we forget, or never knew, how it is instantiated in English in particular. For the latter, we have to observe and integrate the evidence that is provided in the surface forms of the exposure language.

If I had to summarize the odyssey that this succession of studies chronicles, the two-line conclusion might be: Language acquisition is obviously impossible, but somehow they do it. Empiricism is innate, but somehow we get over it.

At peril of going on forever here, I do want to say a few more words about how this enterprise unfolded, for it was certainly a product of "groupthink." Henry Gleitman (ultimately the teacher of all of us collaborators) started a research seminar in which we all continued to talk for almost half a century. After a while, it came to be called "the cheese seminar," because of the refreshments provided halfway through these long evenings of talk. And then, with the addition of insights and methods from Trueswell's students joining this larger group, "the cheese and trackers seminar"—not only cheese and not only eye-tracking. It was decades of cross talk and collaboration, continuing to this very day, for a great circle of young scientists, some of whom I've mentioned in sketching the styles of investigation and explanation that emerged from this group. And I'm glad to say, more than proud, that this seminar continues: I am particularly proud to say that in the present incarnation of this kind of research seminar, now led by Trueswell, my very first postdoc and early collaborator, Sharon Armstrong (professor of psychology at LaSalle University), continues to attend and to contribute to the next generation of this half-century-long conversation.

Language learning isn't really instantaneous; it happens in real time.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was partially funded by a grant to Lila R. Gleitman and John C. Trueswell from the National Institutes of Health (1-R01-HD-37507).

LITERATURE CITED

Armstrong SL, Gleitman LR, Gleitman H. 1983. What some concepts might not be. *Cognition* 13:263–308

Brown R, Cazden C, Bellugi U. 1968. The child's grammar from I to III. In *Minneapolis Symposium on Child Psychology*, vol. 2, ed. JP Hill, pp. 28–73. Minneapolis: Univ. Minn. Press

Carey S. 1982. Semantic development: the state of the art. In *Language Acquisition: The State of the Art*, ed. E Wanner, LR Gleitman, pp. 347–89. Cambridge, UK: Cambridge Univ. Press

Cartmill EA, Armstrong BF, Gleitman LR, Goldin-Meadow S, Medina TN, Trueswell JC. 2013. Quality of early parent input predicts child vocabulary 3 years later. *PNAS* 110:11278–83

Chomsky N. 1957. *Syntactic Structures*. The Hague: Mouton

Chomsky N. 1959. A review of B.F. Skinner's *Verbal Behavior*. *Language* 35:26–58
Chomsky N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press
Connolly AC, Fodor J, Gleitman LR, Gleitman H. 2007. Why stereotypes don't even make good defaults. *Cognition* 103:1–22
Feldman H, Goldin-Meadow S, Gleitman LR. 1978. Beyond Herodotus: the creation of language by linguistically deprived deaf children. In *Action, Symbol, and Gesture: The Emergence of Language*, ed. A Lock, pp. 351–414. New York: Academic
Fisher C, Gleitman H, Gleitman LR. 1991. On the semantic content of subcategorization frames. *Cogn. Psychol.* 23:331–92
Gillette J, Gleitman H, Gleitman LR, Lederer A. 1999. Human simulations of vocabulary learning. *Cognition* 73:135–76
Gleitman H, Gleitman LR. 1979. Language use and language judgment. In *Individual Differences in Language Ability and Language Behavior*, ed. C Fillmore, D Kemler, W Wang, pp. 103–26. New York: Academic
Gleitman LR, January D, Nappa R, Trueswell JC. 2007. On the give and take between event apprehension and utterance formulation. *J. Mem. Lang.* 57:544–69
Gleitman LR, Rozin P. 1977. The structure and acquisition of reading: relations between orthographies and the structure of language. In *Toward a Psychology of Reading*, ed. A Reber, D Scarborough, pp. 55–141. Hillsdale, NJ: Erlbaum
Harris ZS. 1951. *Methods in Structural Linguistics*. Chicago: Univ. Chicago Press
Harris ZS. 1952. Discourse analysis. *Language* 28:1–30
Hirsh-Pasek K, Gleitman LR, Gleitman H. 1978. What did the brain say to the mind? A study of the detection and report of ambiguity by young children. In *The Child's Conception of Language*, ed. A Sinclair, RJ Jarvella, WJM Levelt, pp. 97–132. Berlin: Springer
Hirsh-Pasek K, Treiman R. 1982. Doggerel: motherese in a new context. *J. Child Lang.* 9:229–37
Hume D. 1738–1740. *A Treatise of Human Nature*. Oxford, UK: Oxford Univ. Press
Klima ES, Bellugi U. 1979. *The Signs of Language*. Cambridge, MA: Harvard Univ. Press
Landau B, Gleitman LR. 1985. *Language and Experience: Evidence from the Blind Child*. Cambridge, MA: Harvard Univ. Press
Levinson SC. 1996. Frames of reference and Molyneux's question: cross-linguistic evidence. In *Language and Space: Language, Speech, and Communication*, ed. P Bloom, MA Peterson, L Nadel, MF Garrett, pp. 385–436. Cambridge, MA: MIT Press
Li P, Abarbanell L, Gleitman L, Papafragou A. 2011. Spatial reasoning in Tenejapan Mayans. *Cognition* 120:33–53
Li P, Gleitman LR. 2002. Turning the tables: language and spatial reasoning. *Cognition* 83:265–94
Lidz J, Gleitman H, Gleitman L. 2003. Understanding how input matters: verb learning and the footprint of universal grammar. *Cognition* 87:151–78
Locke J. 1690. *An Essay Concerning Humane Understanding*. London: Thomas Bassett
Medina TN, Snedeker J, Trueswell JC, Gleitman LR. 2011. How words can and cannot be learned by observation. *PNAS* 108:9014–19
Naigles L. 1990. Children use syntax to learn verb meanings. *J. Child Lang.* 17:357–74
Naigles L, Gleitman H, Gleitman L. 1993. Syntactic bootstrapping and verb acquisition. In *Language and Cognition: A Developmental Perspective*, ed. E Dromi, pp. 104–40. Norwood, NJ: Ablex
Nappa R, Wessell A, McEldoon KL, Gleitman LR, Trueswell JC. 2009. Use of speaker's gaze and syntax in verb learning. *Lang. Learn. Dev.* 5:203–34
Newport EL. 1975. *Motherese: the speech of mothers to young children*. PhD thesis, Univ. Pa., Philadelphia
Newport EL, Gleitman H, Gleitman LR. 1977. Mother, I'd rather do it myself: some effects and non-effects of maternal speech style. In *Talking to Children*, ed. CE Snow, CA Ferguson, pp. 109–49. Cambridge, UK: Cambridge Univ. Press
Papafragou A, Cassidy K, Gleitman LR. 2007. When we think about thinking: the acquisition of belief verbs. *Cognition* 105:125–65
Pinker S. 1984. *Language Learnability and Language Development*. Cogn. Sci. Ser. 7. Cambridge, MA: Harvard Univ. Press

Plato. 2009 (ca. 385 BCE). *Phaedo*, transl. D Gallop. Oxford, UK: Oxford Univ. Press
Rodd LJ, Braine MDS. 1971. Children's imitations of syntactic constructions as a measure of linguistic competence. *J. Verbal Learn. Verbal Behav.* 10:430–43
Senghas A, Coppola M, Newport EL, Supalla T. 1997. Argument structure in Nicaraguan Sign Language: the emergence of grammatical devices. In *Proceedings of the 21st Annual Boston University Conference on Language Development*, vol. 2, ed. E Hughes, M Hughes, A Greenhill, pp. 550–61. Somerville, MA: Cascadilla
Shipley E, Smith C, Gleitman L. 1969. A study in the acquisition of language: free responses to commands. *Language* 45:322–42
Smith LB, Yu C, Pereira AF. 2011. Not your mother's view: the dynamics of toddler visual experience. *Dev. Sci.* 14:9–17
Trueswell JC, Medina T, Hafri A, Gleitman L. 2013. Propose but verify: Fast mapping meets cross-situational word learning. *Cogn. Psychol.* 66:126–56
Whorf BL. 1956. *Language, Thought, and Reality: Selected Writings*. Cambridge, MA: MIT Press
Woodard K, Gleitman L, Trueswell J. 2016. Two- and three-year-olds track a single meaning during word learning: evidence for propose-but-verify. *Lang. Learn. Dev.* 12:252–61
Yu C, Smith LB. 2007. Rapid word learning under uncertainty via cross-situational statistics. *Psychol. Sci.* 18:414–20

Annual Review
of Linguistics

Volume 5, 2019

Contents

Errata

An online log of corrections to *Annual Review of Linguistics* articles may be found at http://www.annualreviews.org/errata/linguistics