CENTER FOR RESEARCH IN LANGUAGE

April 1989 **Vol. 3, No. 4**

The monthly newsletter of the Center for Research in Language, University of California, San Diego, La Jolla CA 92093. *(619) 534-2536*; electronic mail: *crl@amos.ling.ucsd.edu*

• • •

CONTENTS

EDITOR'S NOTE

This newsletter is produced and distributed by the **CENTER FOR RESEARCH IN LANGUAGE**, a research center at the University of California, San Diego, which unites the efforts of researchers in such disciplines as Linguistics, Cognitive Science, Psychology, Computer Science, Communication, Sociology, and Philosophy, all of whom share an interest in language. We regularly feature papers related to language and cognition (1 - 10 pages, sent via email) and welcome response from friends and colleagues at UCSD as well as other institutions. Please forward correspondence to:

Teenie Matlock, Editor
Center for Research in Language, C-008
University of California, San Diego 92093
Telephone: (619) 534-2536
Email: *crl@amos.ling.ucsd.edu*

SUBSCRIPTION INFORMATION

If you are currently receiving a hardcopy of the newsletter only and have an email address to which the newsletter can be sent, please forward that information to CRL.

Can we send you the email version only to save printing and mailing costs? If you require a hardcopy in addition, please request it and we will be happy to send you one.

If you require the unformatted nroff version of the newsletter, please request it from CRL after you have received the current regular formatted version.

If you know of others who would be interested in receiving the newsletter, please forward the email or postal mailing address to CRL. Thank you.

BACK ISSUES

Back issues of this newsletter are available from **CRL** in hard copy as well as soft copy form. Papers featured in previous issues include the following:

The Cognitive Perspective
Ronald W. Langacker
Department of Linguistics, UCSD
vol. 1, no. 3, February 1987

Toward Connectionist Semantics
Garrison Cottrell
Institute for Cognitive Science, UCSD
vol. 1, no. 4, May 1987

Dimensions of Ambiguity
Peter Norvig
Computer Science, UC Berkeley
vol. 1, no. 6, July 1987

Where is Chomsky's Bottleneck?
S.-Y. Kuroda
Department of Linguistics, UCSD
vol. 1, no. 7, September 1987
(2nd printing of paper in no. 5, vol. 1)

Transitivity and the Lexicon
Sally Rice
Department of Linguistics, UCSD
vol. 2, no. 2, December 1987

Formal Semantics, Pragmatics, and Situated Meaning
Aaron Cicourel
Department of Sociology, UCSD
vol. 2, no. 3, January 1988

Rules and Regularities in the Acquisition of the English Past Tense
Virginia Marchman
Department of Psychology, UC Berkeley
vol. 2, no. 4, April 1988

A Geometric Conception of Grammar
S.-Y. Kuroda
Department of Linguistics, UCSD
vol. 2, no. 5, June 1988

Harris and the Reality of Language
S.-Y. Kuroda
Department of Linguistics, UCSD
vol. 3, no. 1, September 1988

A Connectionist Perspective on Prosodic Structure
Mary Hare, Dept. of Linguistics, UCSD
David Corina, Dept. of Psychology, UCSD
G.W. Cottrell, Dept. of Computer Science, UCSD
vol. 3, no. 2, November, 1988

ANNOUNCEMENTS

Cognitive Linguistics Workshop

Date: May 13th and 14th, 1989
Location: University of California, San Diego
Topics: Recent linguistic work relating to cognitive linguistics
Speakers: Ronald Langacker, George Lakoff, Eve Sweetser,
Gilles Fauconnier, David Zubin, and others, including
UCSD and UCB students

For further information contact:

Ronald Langacker, Chair
UCSD Department of Linguistics, C-008
La Jolla, CA 92093
email: rwl@ bend.ucsd.edu
phone: (619)-534-3600

****

CENTER FOR RESEARCH IN LANGUAGE Annual Report

The **CRL** 1987-88 Annual Report is ready. Please forward requests to:

Center for Research in Language, C-008
University of California, San Diego
La Jolla, CA 92093
or yvonne@ amos.ucsd.edu

****

Recent Talks

On April 12, 1989, Richard Durbin of Stanford University Psychology Department gave a talk entitled "New Types of Computational Units for Learning Networks".

On April 10, 1989, Nils Nilsson of Stanford University Computer Science Department gave a talk entitled "Action Networks".

The following Technical Report (# 8902) is available from **CRL**:

Pattern Association in a Back Propagation Network: Implications for Child Language Acquisition

Kim Plunkett
University of Aarhus, Denmark

Virginia Marchman
University of California, San Diego

Abstract

A 3-layer back propagation network is used to implement a pattern association task which learns mappings that are analogous to the present and past tense forms of English verbs, i.e., arbitrary, identity, vowel change, and suffixation mappings. The degree of correspondence between connectionist models of tasks of this type (Rumelhart & McClelland, 1986; 1987) and children's acquisition of inflectional morphology has recently been highlighted in discussions of the general applicability of PDP to the study of human cognition and language (Pinker & Mehler, 1988). In this paper, we attempt to eliminate many of the shortcomings of the R&M work and adopt an empirical, comparative approach to the analysis of learning (i.e., hit rate and error type) in these networks. In all of our simulations, the network is given a constant 'diet' of input stems -- that is, discontinuities are not introduced into the learning set at any point. Four sets of simulations are described in which input conditions (class size and token frequency) and the presence/absence of phonological subregularities are manipulated. First, baseline simulations chart the initial computational constraints of the system and reveal complex "competition effects" when the four verb classes must be learned simultaneously. Next, we explore the nature of these competitions given different type (class sizes) and token frequencies (# of repetitions). Several hypotheses about input to children are tested, from dictionary counts and production corpora. Results suggest that relative class size determines which "default" transformation is employed by the network, as well as the frequency of overgeneralization errors (both "pure" and "blended" overgeneralizations). A third series of simulations manipulates token frequency within a constant class size, searching for the set of token frequencies which results in "adult-like competence" and "child-like" errors across learning. A final series investigates the addition of phonological sub-regularities into the identity and vowel change classes. Phonological cues are clearly exploited by the system, leading to overall improved performance. However, overgeneralizations, U-shaped learning and competition effects continue to be observed in similar conditions. These models establish that input configuration plays a role in determining the types of errors produced by the network - including the conditions under which "rule-like" behavior and "U-shaped" development will and will not emerge. The results are discussed with reference to behavioral data on children's acquisition of the past tense and the validity of drawing conclusions about the acquisition of language from models of this sort.

Please send requests for hard copy to:

Center for Research in Language C-008
University of California, San Diego
La Jolla, CA 92093 *Attn. Yvonne*
or yvonne@ amos.ucsd.edu

Virginia Marchman (marchman@ amos.ucsd.edu)
Kim Plunkett (psykimp@ dkarh02.bitnet)

Competent Scientist Meets the Empiricist Mind

Valerie Walker [1]
Department of Philosophy
University of California, San Diego

Noam Chomsky repeatedly presents the scientific community with arguments designed to show that empiricist theories of the mind are wrong, and that the mind must possess a substantial amount of innate knowledge in order for it to acquire linguistic competence. One of the arguments that he gives is what Ramsey and Stich (1989) label the "Competent Scientist Gambit." Though the Competent Scientist Gambit has held sway over a substantial portion of linguists and developmental psychologists, its soundness appears to have rarely been questioned. This essay proposes to rectify this oversight by examining some of the empirical assumptions that the argument must make in order to work, after explaining how the Competent Scientist Gambit functions.

I

That some constraints are needed at all for language acquisition is uncontroversial. Ever since Goodman adumbrated the problem of projection nearly forty years ago in "The New Riddle of Induction" it has been clear that some system of constraints is necessary for any type of systematic generalization to occur (Goodman,1983). *Abduction,* the inferential projection from some set of data to an hypothesis about what other elements also belong in the set, relies necessarily on methodological principles that cannot be contained in the original data set. These methodological principles are needed to constrain the types of patterns one can use in making generalizations. Without them there is no way to settle upon a pattern that one would use to generate hypotheses. That is, without these principles, there are an indefinite number of projections possible from a set of data, and there is no formal way to chose the "right" projection from aomng the possible projections. The bottom line is that some sort of constraint is necessary for the abductive process of generalization to get underway in the first place.

Children learning a language parallel the abductive process. They project from their pool of primary linguistic data to a set of generalizations that not only allows them to make grammaticality judgments on those data, but also allows them to make grammaticality judgments about sentences not contained in the primary data. Exactly how they accomplish this feat is not known. That there must be some biases concerning the types of generalizations they can make is given. To acquire a language, our learning mechanism must utilize "its innate specifications of certain heuristic procedures and certain built-in constraints on the character of the task to be performed" (Chomsky, 1964, p. 26). What type of constraints these are that allow for the abductive generalizations in language acquisition is the question.

The types of answers to this question fall into two camps: the empiricist and the rationalist. Empiricists claim that the constraints on the mind that allow for the abductive generalizations found in language learning are no different than the constraints on the mind that allow for any type of learning. That is, the empiricist conception of the mind holds that the inherent biases that constitute a brain's learning mechanisms are not domain-specific: "Certain general principles of learning that are common in their essentials to all (or a large class of) organisms suffice to account for the cognitive structures attained by humans, structures that incorporate the principles by which human behavior is planned, organized, and controlled" (Chomsky, 1975a, p. 10). Moreover, the empiricists urge that these general purpose constraints are relatively few in number and relatively simple in their processing -- the complexity of our generalizations arises from recursive iteration of these simple processes. Excluding the constraints provided by the sensory transducers, all abduction then is made possible by "certain analytical data-processing mechanisms or inductive principles of a very elementary sort" (Chomsky, 1975b, p. 47) such that "preliminary analysis of experience is provided by the peripheral processing mechanisms, and one's concepts and knowledge, beyond this, are acquired by application of the available inductive principles to this initially analyzed experience" (ibid., p. 48). Hence, any language that is learned by an empiricist mind must be "relatively independent in its structure of any innate metal faculties" (ibid., p. 51). There are no language-specific mechanisms that ultimately determine the structure of the grammar. Rather, the grammar's structure turns on the type of data inputted to the general purpose inductive devices.

Rationalist conceptions of the mind, on the other hand, purport that the constraints on the mind that provide for language acquisition are domain-specific and relatively complex. That is, beyond the peripheral transducers, "there are innate ideas and principles of various kinds that determine the form of the acquired knowledge in what may be a rather restricted and highly organized way" (ibid., p. 48). So, unlike empiricist conceptions of the mind, the structure of the grammar is ultimately determined by the structure of the constraints; the "general form of a system of knowledge is fixed in advance as a disposition of the mind" (ibid., p. 51). (See also Leiber 1975a, chapter 3, for further discussion of how Chomsky draws the line between the two concepts.)

Chomsky supports the rationalist conception of the mind, that "our systems of belief are those that the mind, as a biological structure, is designed to construct" (Chomsky, 1975a, p. 7), and wants to conclude that "the organism brings, as its contribution to acquisition of a particular language, a highly restrictive characterization of a class of generative systems (potential theories) from which the grammar of its language is selected on the basis of presented linguistic data" (Chomsky, 1965, p. 112). In order to do so, he must demonstrate that any empiricist conception of the mind is untenable. That is, he must show that in principle empiricist minds cannot account for language acquisition. Chomsky uses the Competent Scientist Gambit to try to accomplish this feat. The strategy underlying his argument is to portray a learning mechanism at least as powerful as anything that an empiricist theory of the mind would use, and probably more powerful, and then to show that this mechanism cannot do what all native speakers of a language do. If Chomsky can successfully demonstrate that whatever mechanism a child must have in order to learn a language is more powerful than any mechanisms found in empiricist conceptions of the mind, then he will have shown that all empiricist conceptions must be wrong. The learning mechanism that Chomsky has in mind for this argument is whatever mechanisms a competent scientist brings to bear when devising and testing hypotheses.

Chomsky's argument runs as follows (cf., Chomsky, 1975a, pp. 14-22). We give our scientist a typical set of primary linguistic data from some language. (We can envision hooking a video camera to the top of some unfortunate child's head and letting this camera run for ten years or so. The stacks of video tape that result then would contain all the linguistic data that the child has gotten over the course of his first ten years.) The task before the scientist is to discover the projections that the child exposed to the data ends up using as an adult speaker. Children succeed in constructing grammars that project well beyond their primary linguistic data. We demand exactly the same results from the scientist and expect her to project from the data given her to "correct" grammar; that is, we expect her to project to the grammar that correctly describes the behavior of speakers of the language from which she was given data. Chomsky concludes that, given only the stacks of video tape or whatever and science's best analytical techniques, the competent scientist cannot ever do what children do all the time. He concludes that "knowledge of grammatical structure cannot arise by application of step-by-step inductive operations ... of any sort that have yet been developed within linguistics, psychology, or philosophy" (Chomsky, 1965, p. 57). Our scientist will not be able to discover the grammar that our child learns when he acquires his language.

Chomsky is not claiming that the scientist could not figure out and formalize the grammar the child has. He is willing to assume that if anyone can, our competent scientist can. The fact that no one has really been able to do it yet is not his point. Rather, Chomsky's point is that when trying to chose the correct grammar -- the one the child actually acquires -- from all the grammars that fit the primary linguistic data, the scientist would become hopelessly ensnared in Goodman's projection problem. To assume that the competent scientist could come up with the correct grammar is also to assume that she could come up with an endless number of wrong grammars, grammars that fit the primary linguistic data, but do not project in the same manner as the child's to sentences not in the primary set. "All concrete attempts to formulate an empirically adequate linguistic theory certainly leave ample room for mutually inconsistent grammars, all compatible with primary data of any conceivable sort" (Chomsky, 1975b, p.37). Whatever scientific techniques the scientist brings to bear in choosing among the possible grammars to come up with the grammar of the language from which the primary data were actually drawn will not be sufficient to narrow the possible choices to a unique one, for "the language each person acquires is a rich and complex construction hopelessly underdetermined by the fragmentary evidence available" (Chomsky, 1975a, p. 11). Because Goodman showed that there are an indefinite number of projections possible from a set of data to generalizations about those data and that there is no formal way to chose the most simple and elegant generalization from the indefinite number of generalizations, scientific techniques alone cannot overcome problem of projection -- constraints different from general scientific methodology are required to go from the primary linguistic data to a unique grammar. The problem with the Competent Scientist Gambit is that the competent scientist could come up with an indefinite number of grammars that all account for the data she is given, that diverge in judgments made about data not found in the primary set, and (at least some of them) be comparably simple, parsimonious, consilliant, etc. The competent scientist "is frustrated by limitations of available evidence faced by far too many possible explanatory theories, mutually inconsistent but adequate to the data" (Chomsky, 1975a, p. 11). No final choice of the "right" grammar would be possible.

Thus, insofar as it is plausible to view the competent scientist as at least as powerful as the empiricist mind, if the competent scientist cannot perform the child's task, then no empiricist learning mechanism could either. The empiricist conception of language acquisition must be wrong, and some type of rationalist constraints are necessary. In order "to explain how a rich and highly specific grammar is developed on the basis of limited data that is [sic] consistent with a vast number of other conflicting grammars" (Chomsky, 1972, p. 174) one must posit some "innate human faculte de langage " (Chomsky, 1975a, p. 37).

II

The Competent Scientist Gambit maintains that empiricist conceptions of the mind cannot explain language acquisition because a scientist cannot predict the grammaticality judgments that a child will make on the basis of that child's primary linguistic data and the scientific method alone. Without additional help, the competent scientist cannot overcome Goodman's problem of projection, and some sort of further constraints or biases are needed in order to narrow the possible projections down to a single set that gives the same grammaticality judgment as the child on any arbitrarily chosen sentence.

However, as stated above, the Competent Scientist Gambit implies that empiricist conceptions of learning are inadequate only if the competent scientist represents constraints at least as strong as any found on any empiricist mind. In other words, Chomsky must assume in the Competent Scientist Gambit that whatever constraints and biases a scientist has, they are comparable to the relatively simple and relatively general constraints that an empiricist conception of the mind engenders. Here we find the first empirical challenge, for this assumption may not be warranted.

That the competent scientist and any version of the empiricist mind do not have the same constraints is obvious. The question raised by the challenge though is whether these different constraints are comparable in the relevant way such that whatever an empiricist mind can do, the competent scientist can do too. There appears to be at least two ways in which a competent scientist and an empiricist learning device may relevantly diverge. First, the competent scientist is denied any interaction with the data, while the child presumably can interact as much as he likes. Chomsky does not believe this difference to be significant. He assumes that we can successfully model as an instantaneous process any type of inductive learning that occurs. All relevant aspects of language learning can be captured in the following "instantaneous" process. A set of primary linguistic data confronts the child as a unit. In response to these data, the child generates all possible inductive hypotheses that are equally simple, and account for and project from the set. These inductive hypotheses form his set of candidate grammars. Then, innate constraints on the types of inductive generalizations he is allowed to make spring into action and chose the correct grammar from the set of candidates. Chomsky assumes that the generation of hypotheses and the choice of the correct one from the candidates take place in a single time step; that is, all the processes involved in learning a language occur instantaneously.

Chomsky claims that disregarding time when pondering the problems of acquisition is merely a move that simplifies the task and has no impact on the types of theories we could create to account for grammatical projections. However, no where in Chomsky have I found a place in which he gives principled reasons for accepting this claim. As far as I can see, the claim that the simplifying assumption of instantaneousness introduces no complications into accounts of language acquisition goes unargued.

In fact, Morgan (1986) claims the opposite, that this assumption by Chomsky forces one to introduce unnecessary innate constraints to account for learning. He believes that the instantaneous model of learning unnecessarily complicates the story that eventually must be told of how children actually acquire. If a child is allowed to interact with his environment when creating hypotheses to account for data, then the interaction itself acts a constraint on learning. (See also Wexler and Culicover 1980, pp. 94-97, for additional discussion.) He believes, as do Wexler and Culicover, that a non-instantaneous generate-then-test model is a methodologically superior way to understand the problem of projection in linguistics than is Chomsky's instantaneous generate-and-chose model. Unfortunately, I cannot immediately see how a non-instantaneous generate-then-test model requires fewer constraints on learning than does Chomsky's instantaneous generate-and-choose model, and Morgan does not deign to spell out the link. Certainly, a non-instantaneous model is psychologically more realistic, but since Chomsky makes no claim that the instantaneous model is psychologically plausible, realism gives no obvious advantage to non-instantaneous accounts. What one must show is that the interaction between learner and environment via the child's hypothesis testing when there is a failure in comprehension decreases the total number of constraints needed to acquire a grammar. Otherwise, the assumption of instantaneousness has no effect on question of whether the types of biases required to narrow to the search space sufficiently in order to settle on a single grammar are rationalist.

One might think that the child may require fewer constraints than the competent scientist because the child could devise tests for his hypotheses in a way that the competent scientist could not. This increased ability to test decreases the number of constraints needed because the child's actual test would replace a constraint the competent scientist needs in order to make the same projections. For example, the child could entertain two or more grammars that account for his input thus far, but that differ in their projections. This child could then utter a sentence that one grammar considers grammatical but the remaining grammars do not. This utterance would then constitute the test differentiating the hypothesized grammars, and depending upon how the child's "subjects" react to his utterance, he might now be able to exclude some grammars from being his guess at the grammar underlying the language he is trying to learn. Obviously, the competent scientist cannot test her hypotheses in this manner -- there is no one on which she could try out possible utterances. Prima facie, it seems then that the child's interaction with his environment allows him to make choices that the scientist has no way of making, and if the scientist is to make the same choices, she must be additionally constrained. However, this appearance is misleading. Remember that the scientist gets exactly the child's primary linguistic data. If she is sufficiently clevver (we assume that she is), then whatever hypothesis the child dreams up, the scientist will to. Of course, of the child's testing his hypotheses will be in the data that the scientist is given. So, like the child, the scientist can eliminate the hypotheses that turned out to be erroneous when tested because the test is in the data. All the scientist has to do is look for it. It is true that the scientist cannot test hypotheses that fit the data and that the child did not test. And if the scientist is smarter than the child she will have a bigger pool of hypotheses to chose from in the end, but this point does not show that interactive models require fewer constraints to narrow the set of possible grammars to one. Rather, it shows that the more ingenious you are, the more constraints you need to accomplish this task. However, that the scientist is smarter than the child does not show that the interactive account entails fewer constraints than an instantaneous one per se. It appears that whatever the child can do in his interactions with the environment, the scientist can duplicate in her examination of the set of data. Furthermore, she can duplicate this feat without using more constraints than the child.

Even if empiricist conceptions of learning entail non-instantaneous hypothesis generation and testing while the Competent Scientist Gambit remains restricted to instantaneous procedures, such a difference as of yet implies no divergence in results given the same constraints on both stories. This difference between empiricist minds and competent scientists does not appear to be significant for our purposes since both seem to require the same constraints in order to settle upon a unique grammar. Thus, insofar as we use the Competent Scientists Gambit to discover the types of biases necessary to acquire language and the instantaneous assumption has no effect on this issue, then this disanalogy between the mind and the scientist remains uninteresting.

A second way in which the competent scientist and an empiricist mind may diverge is in their respective criterion for chosing a grammar. Scientific principles may not correspond to whatever methodology a child uses in choosing his grammar. Chomsky recognizes this disanalogy, and remarks simply that "that poorly understood property of theories that leads a scientist to select one rather than another" is not the same as the "weighting function" that a child uses to "[select] a grammar ... over others that also ... are compatible with the data" (Chomsky, 1972, p. 289). (See also Chomsky 1984, pp. 170-171, and 1975a, section 3.) That the criterion of "simplicity" is not identical with a child's "weighting function" is obvious. The question is whether this difference is relevant given our gambit. Are there any instances of an empiricist theory that can accomplish what a competent scientist could not? Until recently, the answer appeared to be no. However, with the advent of connectionism and parallel distributed processing models of learning, the answer may change.

"Connectionism" refers to a new type of cognitive modeling that uses networks of interconnected units to process information in parallel. (See Rumelhart, et al. 1986 for discussion.) The idea underlying connectionist models is that complex information processing can emerge from the interactions of large numbers of simple processing nodes. Each node is structurally identical to all the others, and each takes concurrent incoming signals and uses them to compute the value of its output. Once the computation is complete, these nodes either "excite" or "inhibit" their neighboring nodes by sending a positive or negative signal up the connecting pathways. The paths that connect the nodes have different "weights" whose values specify the strength of the connections between each pair of nodes. The system is parallel in that many nodes carry out their computations and output their signals at the same time. The weights interact with the output signals traveling along the connection via some propogation rule, which determines the exact value of the signal that inputs to the next nodes in the network. So, input to nodes is a function of the connection's weight strength and the value of the signal traveling along the connection, and the subsequent output of nodes is a function of the input and the transformation rule of the node.

One way to alter the way information is processed in the network is to change the connection weights between the nodes. If the weights are changed appropriately in response to an error signal that is sent back along the connections, then the network exhibits a rudimentary form of learning. These fairly simple learning strategies allow the networks to "program" themselves to perform some task. Typically, a network begins with arbitrary weights set on its connections. It then undergoes a training period in which it is presented with sets of stimuli input to which it produces some output in response. The actual output response is compared with the "target" output that correctly weighted networks should give. An error signal is computed as a function of the difference between actual response and target response. This error signal is then propogated back through the network, adjusting the weights of the connections, according to some computational learning algorithm. When the computed error reaches a certain minimum, then the network has learned whatever task it was being trained to perform. This network can now make educated guesses about how to process input not included in the training set but relevant for the problem that the network is trained to handle. The network learned successfully if it can generalize appropriately to a superset of data.

I am spelling out the details of connectionist networks to impress upon the reader how "empiricist" these models are. If anything could be an empiricist model of the mind, then connectionist models should be. Complex information processing arises from the parallel interaction of these relatively "stupid" nodes, who are structurally uniform and execute very simple computations. The learning algorithm used to train networks is general purpose and has been used to train networks to do a number of diverse cognitive tasks, from recognizing the sonar echoes of mines to visual facial recognition. Whatever constraints exist in the architecture are certainly relatively few in number and simple in their processing. All connectionist models of this type appear to meet Chomsky's criterion that empiricist learning machines posses only "data-processing machines ... of a very elementary sort" Chomsky, 1975b, p. 47).

One example of complex results from a simple connectionist machine relevant to the problem of language acquisition is Hanson's and Kegl's PARSNIP (Hanson and Kegl 1987). (For other examples of what a connectionist language processing devices can do, see Fanty 1985, Cottrell 1985, Waltz and Pollack 1985, Selman and Hirst 1985, Rumelhart et al. 1986, Charniak and Santos 1986, and Elman 1988.) This network is an auto-associator, i.e. it reproduces whatever it receives as input as its output. When it was trained on sets of syntactically tagged natural language sentences, the researchers discovered that this network, along with producing the correct output, could also produce "grammar-like behavior" while performing various language-like tasks. It learned the correct syntactic category names for the position of sentences on the training set and was able to successfully generalize to 1,000 sentences not in the training set and presented to the system after it had been trained. It can complete sentences, and can even recognize novel sentence patterns that were not presented in the original data set. Finally, like human language users, it "correctly reproduces test sentences reflecting deep center-embedded patterns which it has never seen before while failing to produce multiply center-embedded patterns" (Hanson and Kegl 1987, p. 106).

Granted, PARSNIP does embody certain implausible psychological features, such as having complete sentences being presented in a single time step. However, Hanson and Kegl point out that

> there are important parallels between the task given to PARSNIP and the task that arises for children as they learn a natural language. Both PARSNIP and the child are only exposed to sentences from natural language, they both must induce general rules and larger constituents from just the regularities to which they are exposed, both on the basis of only positive evidence. PARSNIP's ability to generalize knowledge of constituent structure has been extracted from its experience with natural language sentences. (ibid., p. 117)

What Hanson and Kegl have shown is that for the linguistic tasks that PARSNIP can perform, the architectural constraints needed to do those tasks can be empiricist. An empiricist mind could learn to make whatever abductive generalizations PARSNIP can make. It could learn a grammar at least as well as PARSNIP does.

Of course, we do not yet know whether a connectionist machine can really learn a language, but it may not seem so unreasonable now to entertain the thought that it very well might be capable of some sort of sophisticated language acquisition. If a connectionist network could learn to project from primary linguistic data to beyond the data set as the child does and if a competent scientist cannot discover the "correct" projection algorithm, then we have indeed found a significant difference between the capabilities of an empiricist mind and a rational scientist. Of course, the importance of this divergence does depend on the actual limitations of connectionist architectures, which presently remain unknown. However, what the advent of connectionist models like PARSNIP and others do demonstrate is that the assumption that the competent scientist is at least as strongly biased as an empiricist mind may be unwarranted. At least, this premise requires careful argumentation to show that competent scientists approximate empiricist minds in the relevant ways. So far, this argumentation has not been forthcoming.

Is there a way to reformulate the Competent Scientist Gambit so that it takes possible significant divergences between the competent scientist and empiricist minds into account? The short answer appears to be no, not in a way that would make Chomsky happy. Chomsky's project is to provide an argument showing that in principle empiricist conceptions of the mind cannot account for language acquisition. However, given the possibility that the scientist is not biased in a manner that approximates the biases present in all empiricist conceptions of the mind, the burden falls to Chomsky to show that this possibility does not obtain. Unfortunately, there are no accepted empiricist models of the mind, nor is it clear what the biases must be in any empiricist conception. Chomsky can't show that he has biased his competent scientist in the way comparable to the way empiricist minds are biased because no one knows what the biases are. He loses his principled argument, for it now hangs upon exactly which constraints one needs in an empiricist model of the mind. Instead of relying solely on Goodman's problem of projection, his argument also turns on contingent (and, as of yet, unknown) matters of fact.

III

The only way I see to salvage Chomsky's general project of determining what types of constraints are necessary for language acquisition is to constrain the scientist (or some other learning machine) by fiat such that the scientist has exactly the constraints that some empiricist theory claims exist on the mind. We can then examine the power of the constrained scientist for each empirical individually and determine for each theory whether it is powerful enough to account for the abductive generalizations in linguistic projections. Let us suppose then that we know whatever biases are supposedly present in some accepted empirical conception of the mind and that we know how to translate these constraints onto a model that learns to make projections. We can now pose the following question: Could our model learn to project exactly as the child projects, given the same set of primary linguistic data that the child receives? That is, could our model do what the child does, and learn his language's grammar? If the answer is no, then we can argue that at least this empiricist conception of the mind is not strong enough to account for language acquisition -- either a different empiricist conception or some rationalist constraints are needed. We argue this because we assume that children all learn to project in the same way, especially if they receive exactly the same primary linguistic data. We generalize from this assumption and claim that if the competent scientist cannot discover the projections a child actually makes, then the constraints on the competent scientist do not allow her to discover the projections of the grammar of the language that the child speaks. However, that all speakers of the same language project in the same way is an empirical assumption that Chomsky at least does not attempt to support. Herein lies the second empirical charge, that it is in fact false that all speakers of a language project in exactly the same way. If this be the case, then whether the competent scientist fails to produce the same projections as the child is immaterial. No Competent Scientist Gambit would tell us anything about the types of constraints needed in order to learn a language.

Chomsky is very explicit about assuming that the same projections are "attained virtually uniformly [across] the population" (Chomsky 1984, p. 33) and that this assumption is necessary for the Competent Scientist Gambit to work. (See also Chomsky 1965 and 1972 for discussion of this "assumption.") However, he appears to view this assumption as more a statement of unproblematic fact than a premise that requires justification. Is he correct to make such a leap of faith? The evidence suggests not. Gleitman and Gleitman (1970) investigate whether everyone in the same linguistic community projects in exactly the same way by testing how people in a homogeneous community paraphrase compound nouns. In a nutshell, their conclusion is that "the results cast considerable doubt on some of the equalitarian assertions of the transformational linguists" (Gleitman and Gleitman 1970, p. 105).

It is trivially true that all competent speakers of English can paraphrase at least some simple instances of compound nouns, e.g. *dog-house*, *work-sheet*. It is probably true that many less than competent speakers of English can also paraphrase such examples. However, the forms found in these simple instances can be iterated to produce more complex constructions. For example,

> the *egg-plant* is the plant shaped like an egg; and the *egg-plant plant* is the plant on which the egg-plant grows; and the *egg-plant-plant plant* is the place where they make egg-plant plants; and the *egg-plant-plant-plant plant* is the spy in the egg-plant-plant plant, and so forth. (ibid., p. 102)

Gleitman and Gleitman ask whether and how competent speakers unpeel these more complicated constructions.

To test this question, they devised 144 different compounds that were novel (for the most part), involved two applications of compounding rules, were equal in length (both in number of words and of syllables), were composed of simple vocabulary items, and were fairly short. They randomly ordered the stimuli into blocks of twelve, and then presented them to twelve subjects, all of whom were monolingual, white, urban, female English speakers from the mid-Atlantic. The subjects were asked to paraphrase the meaning of each compound noun.

What "stands out immediately" when looking at the results of the experiment is that there is "an overwhelming difference" in the answers given by the participants (ibid., p. 115). Furthermore, not only did the subjects differ in the absolute number of "errors" made, but they also differed in the type of errors they committed. Some subjects made very few errors (the least number made was five), while other subjects made a lot (the most reported was 108). Some subjects make errors reflecting word order; others made errors reflecting word stress. (Here an "error" means an answer that deviates from the response that Gleitman and Gleitman predicted.) These results imply "qualitative differences among the [subjects] ... in the way they approached the paraphrasing task" (ibid., p. 118). Simply put, different subjects paraphrased differently.

In order to determine whether the "errors" were systematic, Gleitman and Gleitman ran a second experiment in which the subjects were asked to chose the best paraphrase from a choice of two for each compound noun. The choices were based on the syntactic structure of the previous responses, although rarely were a subject's exact answers available. Again, the results are striking. The "accuracy on the forced-choice task is greater when the correct alternative is the subject's own paraphrase than when it is a paraphrase the subject did not herself devise." Moreover, the situation exactly reverses "for trails containing stimulus items that the subject had paraphrased incorrectly in the free-paraphrase task: here, choice accuracy was less when the false alternative was the subject's own previous paraphrase, and greater when it was devised by others" (ibid., p. 136). Obviously, the subjects each prefer their own syntactic way of saying things, regardless of whether that way is right or wrong by the Gleitmans' criteria. This conclusion is further strengthened when the Gleitmans tried and failed to teach those who made the most "errors" to perform in a "correct" manner. So far, "no simple familiarization and teaching technique that [they] have been able to devise has resulted in clear learning among the subjects" (ibid., 146) whose parsing differed substantially from what a "correct" grammar would entail.

It would appear that there exists real grammatical differences among speakers of the same dialect. Similar results have also been found in other linguistic tasks (See, for examples, Maclay and Sleator 1960, Pfafflin 1961, Stolz 1967, Levelt 1975a, Smith 1988.) We must now ask: What do these real differences imply for the Competent Scientist Gambit?

At first glance, it may seem that the gambit is doomed. If there is no guarantee that all speakers of one language make the same grammaticality judgments, then there is no way to generalize from what projections a single child makes to the linguistic behavior of an entire community. Furthermore, if a competent scientist comes up with different projections than the child, when both are given exactly the same primary linguistic data, then one can draw no interesting conclusions from that fact, for presumably a parallel result could occur in the linguistic community at large. However, it is not the case that all projections in a given linguistic community from the same data are radically different. Even in the paraphrase experiments discussed above, the subjects were "affected similarly by certain syntactic features of the [noun] compounds" (ibid ., p. 129). That is, some judgments do appear to be constant across all speakers of one dialect. (Obviously, this partial uniformity must be the case if any serious and effective communication is to occur at all.) Gleitman and Gleitman refer to the grammaticality judgments that all speakers of a single dialect must make (excluding random performance errors) as stemming from the "core" grammar, and the type of judgments that are not constant across speakers as coming from a "penumbra" grammar.

So, there appears to be some projections from a set of primary linguistic data that all speakers of a single language must make, and some other projections that are not indicative of any particular language. What is not clear, and what the various formulations of the Competent Scientist Gambit must address, is whether the biases found in an empiricist's mind and the constraints on the solution space are enough to determine all core projections. If all speakers of a language make the same grammaticality judgments about some core set of sentences, then the Competent Scientist Gambit needs to investigate whether the constraints on the empiricist mind are enough to entail this core of grammatical judgments. Unfortunately, there does not appear to be any principled way in which an argument is going to address this question. Whether the biases and constraints are enough is an empirical question and one is going to have to determine exactly what the biases and constraints are, as well as exactly what these limits entail when making projections, before one can begin to seriously address whether empiricist conceptions of the mind allow for the core projections.

Furthermore, because there is no principled way to determine exactly what judgments stem from core projections and which judgements are merely penumbra, the task confronting the Competent Scientist Gambit is even more difficult. If the competent scientist came up with a projection that fits the primary linguistic data but diverged slightly from a core projection, then one could plausibly reply that the competent scientist has not refuted the empiricist conception of the mind; instead all she did was show that the way we had separated the core grammar from the penumbra was wrong. She discovered the real boundary between what all speakers must do and what they can do. In order for a Chomskyan to successfully argue that the empipricist constraints on the mind are not enough to generate the core projection, he must show that the competent scientist can diverge radically from what the core projects. Since we do not know what type of judgments must fall out of the core grammar in order for the Competent Scientist Gambit to work, a Chomskyan would have to show that not only do the empiricist restrictions allow projections different than those the core grammar entails, but also that those projections are very different than what the core entails.

Could the Competent Scientist Gambit show that any empiricist conception of the mind is too weak to discover the projections that the core grammar of a langauge makes? To answer that question requires empirical leg-work. We don't know what any empiricist mind looks like, nor do we know what is included in the core. The nativism issue appears to depend upon purely empirical questions that cannot be decided on the basis of thought-experiments alone. Chomsky's Competent Scientist Gambit is no longer tenable.

1. I would like to thank Steve Stich and Virginia Marchman for their helpful and patient comments and criticisms.

References

Bates, E. and MacWhinney, B. (forthcoming) "Functionalism and the Competition Model" in B. MacWhinney and E. Bates (eds.) The Crosslinguistic Study of Sentence Processing. New York: Cambridge University Press.

Bates, E., Thal, D., and Marchman, V. (1989) "Symbols and Syntax: A Darwinian Approach to Language Development," in N. Krasnegor, D. Rumbaugh, M. Studdert-Kennedy, and R. Schiefelbusch (eds.) The Biological Foundations of Language Development. Oxford: Oxford University Press.

Chomsky, N. (1964) Current Issues in Linguistic Theory. The Hague: Mouton and Company.

___________ (1965) Aspects of the Theory of Syntax. Cambridge, Massachusetts: MIT Press.

___________ (1972) Language and Mind, Enlarged Edition. San Francisco: Harcourt Brace, Jovanovich, Inc.

___________ (1975b) The Logical Structure of Linguistic Theory.

___________ (1975a) Reflections on Language. New York: Random House.

___________ (1984) Modular Approaches to the Study of the Mind. San Diego, California: San Diego State University Press.

Cottrell, G. (1985) "Connectionist Parsing," in Proceedings of the Seventh Annual Cognitive Science Society, pp. 201-211.

Elman, J. (1988) "Finding Structure in Time," CRL Technical Report 8801.

Fanty, M. (1985) "Context-Free Parsing in Connectionist Networks," Technical Report # 174, Department of Computer Science, University of Rochester.

Gleitman, L. and Gleitman, H (1970) Phrase and Paraphrasative Uses of Language. New York: W.W. Norton and Company, Inc.

Goodman, N. (1983) Fact, Fiction, and Forecast -- Fourth Edition. Cambridge, Massachusetts: Harvard University Press.

Hanson, S. and Kegl, J. (1987) "PARSNIP: A Connectionist Network that Learns Natural Language Grammar from Exposure to Natural Language Sentences," in Proceedings of the Ninth Annual Conference of the Cognitive Science Society, 106-119.

Leiber, J. (1975a) Noam Chomsky: A Philosophic Overview. New York: St. Martin’s Press.

Maclay, H. and Sleator, M. (1960) "Responses to Language: Judgments of Grammaticalness," IJAL, 26:275-282.

Morgan, J. L. (1986) From Simple Input to Complex Grammars. Cambridge, Massachusetts: MIT Press.

Peters, S. (1972) "The Projection Problem: How is a Grammar to be Selected?" in S. Peters (ed) Goals of Linguistic Theory. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Pfafflin, S.M. (1961) "Grammatical Judgments of Computer-Generated Word Sequences." Mimeo. Murray Hill, New Jersey: Bell Telephone Laboratories.

Ramsey, W. and Stich, S. (1989) "Connectionism and Three Levels of Nativism," forthcoming.

Rumelhart, D. , McClelland, J., and the PDP Research Group (1986) Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge, Massachusetts: MIT Press.

Selman, B. and Hirst, G. (1985) "A Rule-Based Connectionist Parsing System," in Proceedings of the Seventh Annual Conference of the Cognitive Science Society.

Stolz, W.S. (1967) "A Study of the Ability to Decode Grammatically Novel Sentences," Journal of Verbal Learning and Verbal Behavior. 6:867-73.

Waltz, D.L. and Pollack, J.B. (1985) "Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation," Cognitive Science. 9:51-74.

Wexler, K. and Culicover, P. (1980) Formal Principles of Language Acquisition. Cambridge, Massachusetts: MIT Press.