J. Child Lang. **31** (2004), 963–968. DOI: 10.1017/S0305000904006464 Printed in the United Kingdom

The Multiple Processes Approach: reply to commentaries on 'A multiple process solution to the logical problem of language acquisition' by Brian MacWhinney

These twelve thoughtful commentaries demonstrate interesting shifts in our collective understanding of the 'logical problem of language acquisition.' The bulk of the commentary supports the multiple process approach to the logical problem. At the same time, there is strong and productive disagreement regarding the ways in which conservatism, competition, probabilistic identification, indirect negative evidence, item-based learning, cue construction, monitoring, and constraints make their separate and related contributions to the learning of specific target structures.

Rohde presents a compelling argument in favour of an integrated multiple processes approach. He shows that, without also invoking the notion of probabilistic identification, a simple invocation of competition is not enough to solve the logical problem. The emphasis that Rohde and Chater place on the importance of probabilistic learning stands in contrast to analyses by Pinker and Kirby. Pinker believes that probabilistic learning of any type, including both Bayesian learning and that envisioned by Horning (1969) involves an implicit invocation of innate structure. Kirby follows this same reasoning when he argues that children's Bayesian priors must 'ultimately be determined by their biology.' Of course, neural networks must begin with some small random weights. But these are just random weights, not discrete preordained parameter settings. Universal Grammar is not a theory about how to set initial random weights. So it must be that Pinker and Kirby are instead echoing the general point made by Marcus and Hoff that learning must be based on some initial structure. No one would dispute this. My specific claim is only that this initial structure involves a system for encoding item-based syntactic patterns and that input then drives the elaboration of these patterns.

Rohde believes that a distributed neural network will be able to instantiate the various processes I have described. I find this idea quite attractive. However, the architecture of the required system cannot be just a simple feed-forward net. Instead, it will rely on interlocked resonant pathways (Damasio, 1989), self-organizing maps (Kohonen, 1997), mechanisms for supporting item-based learning (Gupta & MacWhinney, 1997), and systems for attention-driven error-correction (Grossberg, 1987).

Chater and Kirby develop some of the themes we find in Rohde, although in a rather different direction. Chater casts the issue of probabilistic

identification within the framework of minimum description length (MDL) analysis. He proposes a specific example of the application of MDL to language learning. In this example, MDL leads the child to extract final *–s* as a suffix on a present tense verb to mark agreement with a singular subject. Chater notes that addition of this pattern can cut description length in half by generating forms such as *the cows sing* and *the cow sings* while excluding *the cow sing* and *the cows sings*. Moreover, it may well be that what drives initial acquisition of final *–s* is not the need to reduce MDL, but rather the need to parse an incoming sentence in which the verb stem is already stored in the lexicon, but not in the third person. This suggests that we may need to formulate a particular MDL model within the context of a detailed processing mechanism such as item-based learning.

Adopting a view from the study of language evolution, Kirby suggests that the competition between analogy and episodic memory can be viewed as a biological implementation of Bayesian statistics. His specific proposal is that, 'a generalization which is assigned a high prior probability should eventually be rejected by a learner once the likelihood of the observed data has fallen too far.' This is in fact a statement in Bayesian terms of a proposed mechanism for computing indirect negative evidence. My own characterization of this would instead view the biological implementation in terms of the competition between item-based and feature-based pattern learning, as I discuss in further detail below.

The analyses from Hausser and O'Grady can be usefully compared, since they both focus tightly on the status of violations of Structural Dependency such as (1–2).

1. *Is the man who coming is running?*
2. *Is the man who is coming running?*

Following Pullum & Sholz (2002), I have emphasized the extent to which children extract positive evidence to acquire control of the structure in (1). To do this, I argue that they rely on commonly occurring frames such as 'where is the X that …' or 'which is the X that …' These frames provide positive evidence indicating that the fronting of tensed elements marks the questioning the proposition stated in the main clause. This information is stored initially on the syntactic frame of each auxiliary and wh-word (MacWhinney, 1982). Hausser suggests that children treat the attempted movement in (2) as the violation of a pragmatic principle. They will not attempt to raise the verb from a relative clause because it makes no sense to question syntactically frozen (Ross, 1974), pragmatically backgrounded information. After all, he argues, children are concerned not with the enumeration of possible strings, but with learning how to express meanings. I agree with Hausser's analysis and the idea that item-based learning

is particularly sensitive to the identification of the meaningful pragmatic relation between questions and information being questioned.

O'Grady develops the notion of multiple processes in yet another important direction. He notes that the initial *is* in (1) seeks to combine as soon as possible with a subject. It tries to combine at first with *the man*, but the relativizer *who* functions to convert *the man* into a potential complex NP. When the processor has assimilated the string *man who coming*, it then chokes because of the ill-formedness of the relative clause. Lewis & Elman (2001) make this same point, showing how a neural net trained on positive sentences like (2) will choke on (1). This local ill-formedness surely contributes to the ungrammaticality of (1) during comprehension, just as O'Grady and Lewis and Elman argue. However, the results of Crain & Nakayama (1987) are about production. For production, a pragmatic process of the type offered by Hausser seems more important, since the child has access to both foregrounded main clause information and backgrounded embedded clause information at the time of sentence planning. We do not need to adjudicate between these alternative accounts. Rather, it follows from my emphasis on multiple solutions to the logical problem, that item-based learning, pragmatic freezing, and parser failure all play a mutually supportive role in assuring that children avoid sentences like (1), both in comprehension and production.

Li explores the interaction of processes during recovery from an overgeneralization such as *unbuild* for the action of demolishing a tower of Lego blocks. Although it has no head-on competitor, *unbuild* competes less directly with *take apart*, *knock down*, *dismantle*, *demolish* and related forms. At the same time, *unbuild* receives general analogic support from dozens of reversives such as *unfasten* and *unlatch*. The child eventually notices that, unlike its competitors, *unbuild* is never actually attested in the positive input. In this sense, we might say that recovery from production of *unbuild* depends on indirect negative evidence. However, it is even more likely that the child engages in cue construction to rein in *unbuild* by noting that reversives with *un-* require point-by-point reversals of a process. If the child actually were to remove each block from the tower in the order in which it was placed during building, then *unbuild* would be possible. But if the child simply smashes the tower with a blow of the hand, then *demolish* is more appropriate. Thus, we see evidence here for the operation of competition, indirect negative evidence, and cue construction.

Hoff calls attention to an important literature that I was unable to cite, because of space limitations. She shows clearly that the more high quality linguistic input that children receive, the more advanced their language becomes. This is not just some marginal, transitional effect during childhood. Instead, positive evidence has a continuing effect across the lifespan. Despite her emphasis on the importance of input, Hoff expresses reservations

regarding input-driven accounts of language acquisition. For example, she questions the extent to which parents provide fine-tuned input to their children, citing an earlier study by Snow, Perlmann & Nathan (1987). Here it might be useful to examine contrasting findings from Sokolov (1993). Hoff also criticizes a putative reliance of input-driven methods on semantic features. However, there are many input-driven models that rely on non-semantic features such as phonology (MacWhinney, Leinbach, Taraban & McDonald, 1989) or co-occurrence patterns (Farkas & Li, 2001). Finally, Hoff questions the role of indirect negative evidence in recovery from overgeneralization. Here, I am much in agreement with Hoff's intuitions. It strikes me as unlikely that indirect negative evidence would play a major role in early learning. However, I do imagine it plays some role in recovery from errors such as *unbuild* in older children.

Marcus develops a theme that can also be found in the commentaries of Niyogi, Hoff, Pinker, and Kirby. The core of this analysis is that general learning principles are fine, but not enough to 'finally' solve the logical problem. In order to deliver this final solution, we will somehow have to invoke linguistically-specialized machinery. Like Marcus, I agree that there is something special about language. I choose to locate this special gift in the human ability to organize language through item-based patterns. These patterns are not mere lists, as Niyogi seems to think, but rather slot-filler operations that lie at the core of the combinatorials of language. At the same time, I also believe that the ability to engage in item-based learning is uniquely human. However, unlike Hauser, Chomsky & Fitch (2002), I see this ability as emerging phylogenetically from earlier roots in gesture and tool use (MacWhinney, 2002). Moreover, I view sentential recursion as a secondary reflex of the slot-filler operation found in item-based learning.

Given the centrality of item-based patterns in this discussion, it is remarkable that only Naigles took direct issue with their role in language acquisition. She suggests that, although item-based patterns may play some role in early productions, comprehension is driven by highly general constructions, such as the English transitive causative, that are applied in a very liberal fashion. In 1982, I argued that such feature-based constructions are formed through induction from a database of item-based constructions. If a group of items is consistent semantically and syntactically, analogic patterns will quickly emerge, even during the two-word stage. If there are inconsistencies within the group, feature-based pattern extraction will be weaker and slower. Having said this, it seems to me that Naigles' experiments tell us a great deal about the role of emergent feature-based patterns in two-year-olds. Although these patterns are not strong enough to promote reliable generalization in production studies with two-year-olds (Tomasello, 2000), they are strong enough to support statistically significant choice patterns in two-choice and enactment comprehension tasks.

Scholz believes that the multiple solutions approach relies on conservatism alone. This is not correct. It is true that the central role of item-based patterns in my proposed approach bring with it a basic commitment to a certain level of conservatism. However, I am not proposing, as Scholz seems to suggest, that the only mechanism that can move the child beyond conservatism is indirect negative evidence. On the contrary, as I proposed in 1982, I believe that feature-based patterns are extracted on the basis of item-based patterns. Once this occurs, the two sets of patterns are in competition. Eventually, in the older child, there may also be additional cue search, monitoring, and computation of indirect negative evidence. However, the basic competition for young children is between item-based patterns and emergent feature-based generalizations. Scholz, Niyogi, and Naigles attempt to characterize the learning I propose as strictly conservative. However, it is actually a competition between conservative item-based learning and liberal, analogic feature-based patterns.

In conclusion, let me say that it was a great privilege to have received such a thoughtful and provocative set of commentaries. My own future work in this area will seek to provide increasingly clear specifications regarding the onset, application, and generalization of item-based patterns. I will, need to correct a few errors in the illustrations I used to characterize the Gold model. It will be important to think about ways in which pragmatics (Hausser), syntactic carpentry (O'Grady), lexical groups (Li), and feature-based constructions (Naigles) contribute information used by multiple processes. It will also be interesting to explore suggestions from Niyogi, Rohde, Chater, Scholz, and Kirby regarding the use of both neural nets and probabilistic models as ways of integrating conservatism and liberalism. Having said this, it seems to me that the principle result of this discussion is the evidence it provides for an ongoing paradigm shift. This shift moves away from a single or dual process account for the logical problem toward a multiple processes account. As Bechtel & Richardson (1993) have shown, shifts of this type are common in the history of science. However, when they occur, it is not enough to simply say that, 'all the factors interact.' Instead, we now have the even more difficult problem of understanding exactly how the processes interact dynamically.

REFERENCES

Bechtel, W. & Richardson, R. C. (1993). *Discovering complexity : decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton University Press.

Crain, S. & Nakayama, M. (1987). Structure dependence in grammar formation. *Language* **63**(3), 522–43.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition* **33**(1–2), 25–62.

Farkas, I. & Li, P. (2001). Modeling the development of lexicon with a growing self-organizing map. In H. J. Caulfield (ed.), Proceedings of the Sixth Joint Conference on Information Science: Association for Intelligent Machinery.

Grossberg, S. (1987). Competitive learning: from interactive activation to adaptive resonance. *Cognitive Science* **11**, 23–63.
Gupta, P. & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: computational and neural bases. *Brain and Language* **59**, 267–333.
Hauser, M., Chomsky, N. & Fitch, T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–79.
Horning, J. J. (1969). *A study of grammatical inference*: Stanford University, Computer Science Department.
Kohonen, T. (1997). *Self-organizing maps*. Berlin: Springer.
Lewis, J. D. & Elman, J. (2001). Learnability and the statistical structure of language: poverty of stimulus arguments revisited. Proceedings of the 26th Annual Boston University Conference on Language Development.
MacWhinney, B. (1982). Basic syntactic processes. In S. Kuczaj (ed.), *Language acquisition: vol. 1. Syntax and semantics*. Hillsdale, NJ: Erlbaum.
MacWhinney, B. (2002). The gradual evolution of language. In B. Malle & T. Givón (eds), *The evolution of language*. Philadelphia: Benjamins.
MacWhinney, B., Leinbach, J., Taraban, R. & McDonald, J. (1989). Language learning: cues or rules? *Journal of Memory and Language* **28**, 255–77.
Pullum, G. & Scholz, B. (2002). Empirical assessment of stimulus poverty arguments. *Linguistic Review* **19**, 9–50.
Ross, J. (1974). Three batons for cognitive psychology. In W. B. Weimer & D. S. Palermo (eds), *Cognition and the symbolic processes*. New York: Wiley.
Sokolov, J. L. (1993). A local contingency analysis of the fine-tuning hypothesis. *Developmental Psychology* **29**, 1008–23.
Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences* **4**, 156–63.