FUNCTIONAL INNATENESS:
EXPLAINING THE CRITICAL PERIOD
FOR LANGUAGE ACQUISITION

James R. Hurford

Department of Linguistics

University of Edinburgh

Professor James R Hurford,
Department of Linguistics,
University of Edinburgh,
Adam Ferguson Building,
40 George Square,
Edinburgh EH8 9LL,
Scotland,
UK.

(Boldface used for emphasis.)

ABSTRACT

In recent years, several explanations have been offered for the critical period in language acquisition, itself, *a priori* a somehwta surprising phenomenon. Two such explanations are considered here. Both studies use computer simulations, but the factors they model are very different.

Hurford (1991) simulates the phylogenetic evolution over hundreds of generations of a species in which the timing of life history traits is under genetic control. The period when an individual is most proficient at language acquisition is just such a life history trait, and is capable of adaptive evolution. Evolutionary simulations lead to a concentration of language acquisition proficiency in the period up to puberty, with a subsequent tailing off.

Elman (1993) demonstrates 'the advantages of starting small' in neural networks learning mini-languages with many of the complex interacting grammatical factors found in real languages. A neural network which starts mature, with a full adult 'working memory' cannot acquire such complex grammatical competence, whereas a net whose attention span is initially limited and then grows with maturation can acquire the appropriate grammar. This explains, in adaptive terms, the existence of a period in which an organism's characteristics, relevant to the language learning task, change, increasing a certain capacity ('working memory') from an immature to an adult value.

These accounts are complementary and mutually compatible. An evolutionary account is proposed, in which genetically controlled 'working memory' size in relation to life history is the variable operated on by natural selection. This account promises to produce a more detailed explanation of the critical period, which can be related to a wider range of data, including the coincidence with puberty and the involvement of sentence processing in language acquisition

The relationships between Elman's 'working memory' and the distinct psychological concept of working memory are also explored.

1 How this paper relates to the formalism/functionalism debate.

The debate between formalists and functionalists cannot be characterized by any absolutely exceptionless generalizations; both species reveal as many subtly distinct phenotypes as there are individual organisms embroiled in the struggle. The Darwinian metaphor is not to suggest that the longest-surviving individuals will 'win' this debate (I hope we're more rational than that!), but to evoke some consciousness of the dimension of phylogenetic evolution, often forgotten in this debate, in preparation for what follows.

With that necessary hedge, one may say that **typically**, if not universally, the debate is between rival explanatory mechanisms which operate within a relatively brief time-span. The innate mechanisms (LAD or UG) postulated by formalists (e.g. Nina Hyams in this volume) apply over a short **ontogenetic** time-span, the few years that it takes a normal child to acquire a first language. The diachronic mechanisms proposed by many functionalists (e.g. by Joan Bybee and Jack DuBois in this volume) operate over somewhat longer periods, during the histories of particular languages, which transcend generations; such 'glossogenetic' change is a subtype of historical-cultural evolution in general.

Neither formalists nor functionalists say much about phylogenetic evolution, the long-term biological evolution of the species. An exception in this volume is Howard Lasnik, who reminds us of a passage in his paper with Chomsky (Chomsky and Lasnik, 1977), in which they allowed that perhaps there could be a valid kind of functional explanation for aspects of language, but that such explanations operated at the phylogenetic level of the evolution of the human species. That passage has seemed to some to be a rather grudging concession to a kind of functionalism that the bulk of functionalists were unlikely to take up, because to take it up would be to concede, on their part, an element of innateness; and innateness, one might conclude, on overhearing the more shrill skirmishes of the debate, is supposed not to be compatible with a functionalist approach.

I placate both your houses. Some characteristics of languages are shaped by diachronic pressures of the kind that Bybee, DuBois, and the grammaticalization theorists envisage, operating over at least several generations. Other characteristics of languages are imposed each generation by children with very specific innate 'ideas' about the grammar they are to acquire. Functionalism and nativism are compatible, in two ways. Firstly, in the short (ontogenetic or glossogenetic) time-scale, they may share the explanatory burden; a language is the eclectic result of the interaction of many factors, functional and innate.

Secondly, and more profoundly, in the long phylogenetic time-scale, functional pressures may actually shape the innate language faculty. This adaptationist idea is resisted by formalists, such as Piattelli-Palmarini (1989) and Chomsky (1968:82-83; 1980:99-100; 1982:29; 1988:167,170) who oppose an adaptationist program in the origins of language. But Pinker and Bloom (1990) have refuted such objections with enough cogency, I believe, to justify pursuing the phylogenetic functionalist-nativist approach. Pinker and Bloom's paper, has, however, been received with only modified rapture by functionalists (e.g. Bates and MacWhinney, 1990) who resist the idea of a domain-specific modular language faculty (e.g. Bates (1994), Bates, Thal and Marchman (1991)). The present paper is neutral on the issue of the modularity or domain-specificity of the language faculty. Other functionalist linguists, such as typologists and grammaticalization theorists, concerned mainly with short-term functional pressures, pay little attention to phylogenetic evolution.

The work described in the present paper brings together a study (Hurford, 1991) perceived by some as supporting a nativist position on language acquisition and a study (Elman, 1993) commonly held as an example of a functionalist (though not Bybeean diachronic) approach. While knee-jerk responses to these papers might assume them to be incompatible, I shall show that they are in fact compatible, and moreover claim that an approach which attempts to produce a synthesis of the leading ideas in both papers promises to produce a more compelling account of the critical period for language acquisition than either paper on its own.

As a final point in this introduction, let me make two pleas for rigor, one applied to theorizing and one applied to the consideration of relevant data.

Computational methods of checking theories, of either formalist or functionalist stripe, add considerable rigor to them. When formal syntactic studies were more formalized than they are now, especially in the heyday of the 'Standard Model' of Transformational Grammar (Chomsky, 1965), computational grammar testing programs were produced. Friedman (1968) was an early example, and Larson et al. (1996) is a recent one, envisaged as a teaching tool, but equally applicable to the mechanical checking out of grammars by a theorist. Syntactic theorists seldom make much use of such programs, which is a pity; any theory which has an axiomatic basis, which claims that certain conclusions follow necessarily from the operation or interaction of certain precisely defined principles, can benefit from the kind of rigorous testing that computational modelling forces. Computational modelling shows theorists the gaps in their theories, and reveals hidden assumptions which they may not always be happy to recognize; and of course it can reveal downright inconsistencies that were simply never imagined, because of the exhaustive and unimaginative way a computer can be made to search a problem space. Human theorists are always too forgiving of theories, especially their own.

Functionalist theories, no less than formal ones, stand to gain by being subjected to computational testing. Where a claim is made that certain factors cause a language, or the language faculty, to take a particular shape, the precise causal mechanisms need to be spelled out explicitly, and their interaction in all possible combinations of relevant circumstances checked. A well known functional explanation in the literature is Hawkins' parsing explanation for certain word order universals (Hawkins, 1990). A problem with such proposals is that one cannot be sure of the most basic requirement; one cannot be sure **that they work**! Computational implementation of the proposed mechanisms and assumed preconditions can show whether such a theory really delivers the goods, under what circumstances, and can shed light on the possible causes of failure. Kirby (1994) has put Hawkins' proposed explanation, and others, to the computational test, leading to a more subtle understanding of its workings. Of course, computer simulation is not a panacea. Computer simulation can be done badly, too; but in general the computer brings a discipline to both formalist and functionalist proposals that is likely to weed out the inconsistent or trivial ones and strengthen the better ones.

Computational simulation and checking of theories may improve their detail and ensure their internal consistency, but it still does not prove them right or wrong. Only confrontation with data can do that. Any mechanism proposed by any theory calls for evidence from real live subjects (normal or abnormal) that the envisaged mechanism actually works in real people. The traditional emphasis within linguistics, which formalists and functionalists inherit, is on model building. Both functionalists and formalists are adept at proposing theoretical explanatory schemes, while leaving real empirical testing to experimental psycholinguists, developmental psychologists, and neuroscientists. Of course there is a place for theoreticians, but they must be sensitive to whatever data can be found relating to their theories. In the final section of this paper, some empirical evidence relating to the theoretical mechanisms deiscussed below, will be mentioned.

2 Two explanations for the critical period

The two studies concerned (Hurford, 1991; Elman, 1993) emerged from very different research communities. While both use computational simulation as a serious tool, and both appeared in the same journal, and both claim to explain the critical period for language acquisition, the papers are in most other respects entirely different in their theoretical backgrounds and concerns. But, I claim, they are complementary and compatible. Table 1 summarizes in schematic form their complementarity.

	Hurford 1991	Elman 1993
Question	When? I.e. timing of critical period	What? I.e. mechanism of critical period
Answer	Puberty	Limited ‘working memory’ in infants
Theoretical framework	Neo-Darwinian evolution	Neural net models of sentence processing
Conclusion	The importance of starting young	The importance of starting small

TABLE 1.

Each study reaches a particular somewhat counterintuitive conclusion. Hurford shows how, within a generally adaptationist theory, an adaptive trait, the capacity to acquire language, may nevertheless decline. And Elman shows how the fact that a cognitive resource is in some sense impoverished can in fact facilitate language acquisition. The next two subsections will briefly describe the two studies.

2.1 Hurford, 1991: why language acquisition capacity declines after puberty.

Central to this study are the notions of **Lifestage** and **Amount of language acquirable**. The critical period phenomenon is represented graphically in a “Language acquisition profile”, a histogram plotting these two factors against each other. Such profiles are assumed to be genetically determined. Given in Figure 1 are (a) a hypothetical, and highly implausible, language acquisition profile; and (b) a language acquisition profile which would show a critical period effect.

.

[*** FIGURE 1 GOES HERE ***]

To explain, a person's lifetime is divided into equal stages (here 10), during which the genes may activate the language acquisition capacity to some degree or other. The degree of language acquisition capacity is quantified as the fraction of a notional "whole language" that can be acquired during a particular lifestage, according to the individual's biologically pre-set lifetime clock. (This biological clock also controls the switches of other life history events, such as puberty, menopause, and various signs of ageing.) The hypothetical profile in Figure 1(a) represents the unlikely phenotype of an individual whose lifetime clock disposes it to acquire 10% of a language at lifestage 2, a further 30% at lifestage 4, 20% at stage 6, and another 10% at stage 9. This individual, therefore, would acquire only 70% of a language during its life, much of it quite late in life. We presume that such an individual would be disadvantaged, relative to an individual possessing the language acquisition profile in Figure (b), in which a critical period effect is represented. The phenotype in Figure 1(b) can acquire 30% of a whole language in the first stage of life, a further 50% in the second stage, 20% in the 3rd lifestage, and another 10% in the 4th. In fact, such an individual has "surplus capacity" (which might be useful in the case of trauma causing loss of language previously acquired).

Language acquisition profile phenotypes such as these were coded, in a computational simulation, by 'genomic' data structures representing strings of alleles, as shown in Figure 2.

[*** FIGURE 2 GOES HERE ***]

To calculate an individual's language acquisition capacity for a given lifestage from such a genetic make-up, add up the numbers for the alleles affecting that lifestage (with the natural proviso that any negative sum counts as zero). Note that both polygenic inheritance (many genes code for one trait) and pleiotropy (one gene codes for many traits) are modelled here, somewhat plausibly approaching the degree of complexity of the actual relation between language traits and genes. The positive and negative values in Figure 2 can be switched at any time by a simulated mutation.

With this apparatus in place, a simulated population of 30 individuals was set up, with 3 individuals at each lifestage, each with a genetic make-up as in Fugure 2. This population was cycled repeatedly through many 'generations', where each generation took the following pattern.

1. Every individual is exposed to language and acquires as much of it as its genetically determined language acquisition profile permits for the lifestage it happens to be at.

2. All individuals are assigned a 'fitness', which is equal to the amount of language they have acquired so far.

3. On the basis of this fitness, parents of the next generation are selected, from among the 'mature' individuals at lifestages 3-10. (That is, puberty is postulated to arrive with lifestage 3.) Those with most language are most likely to be selected as parents.

4. The selected parents breed offspring, who are added to the population with lifestage 1. Sexual reproduction was closely simulated, with newborn individuals inheriting a mix of the genes of their parents, with the concomitant likelihood of similar language acquisition profile phenotypes being propagated.

5. All individuals, except those just born, are aged by one lifestage, with those at lifestage 10 'dying' and being eliminated from the simulation.

6. The cycle starts again at the beginning.

This cycle contains the essential neo-Darwinian elements of genotypic variation, heritability, and selection among phenotypes according to a fitness function.

The results yielded the evolution, under several sets of conditions, of a clear critical period effect, with most language acquisition capacity concentrated in the pre-puberty lifestages, and a sharp decline thereafter. Figure 3 shows some of the results. These are the average language acquisition profiles of the whole evolved population after 1000 generations.

.

[*** FIGURE 3 GOES HERE ***]

An important point about these results is that they do not show the drop-off in language acquisition capacity to be an evolutionary adaptation. Clearly, possessing the ability to acquire language is an advantageous characteristic; losing this capacity potentially loses advantage. But normal humans only acquire their native language(s) once; and once acquired, in the absence of brain injury, language stays. Normally, one does not need to acquire language after puberty, so there is little or no evolutionary selectional pressure to maintain a language acquisition capacity after puberty. In this model (Hurford 1991), the random mutations which create facilitating or inhibiting factors in relation to language acquisition continue, and over time produce mutually cancelling effects in the later lifestages. But language-capacity-inhibiting genes are prevented from invading the parts of the genome relevant to the early lifestages of the phenotype because of the advantages of acquiring language as early as possible. An organism programmed only to begin to acquire language halfway through life will be disadvantaged in relation to an organism which acquires language early. But of two organisms which both acquire language early in life, no great advantage accrues to one which has a language acquisition capacity prolonged into later life, past the time of normal complete acquisition.

In short, the concentration of language acquisition capacity in early life **is** an evolutionary adaptation, and the drop-off in this capacity around puberty is a by-product of its early concentration and, in this model, of attrition by random mutation and genetic recombination affecting the lifestages where there is less or no selectional pressure to maintain it.

In this model, the language possessed by the simulated organisms was reified; language was just some stuff that one could have more or less of, and it was advantageous to get as much as one could in early life. While it is not wrong to speak of 'how much' of a language a learner has acquired (this is what language tests used by language teachers quantify), this simple model does not say anything about exactly **how** the advantage of possessing language is cashed out in more real-life terms.

Fuller details of this model and the simulations can be found in the original paper (Hurford 1991).

2.2 Elman, 1993: the learning advantages of immaturity and resource limitations

Elman tried to train a neural network to learn a language defined by a somewhat complex context free phrase structure grammar. The language differentiated the most basic syntactic categories familiar from English and permitted multiple recursion, embeddings and nested long distance dependencies. Though clearly less complex than the whole of English, the language exemplified some of the central formally problematic aspects of natural languages. Elman's was one of the first programs to actually manage to learn **any** plausibly naturalistic sublanguage. It is notable that formalist theories that put language acquisition at their center have not managed to simulate the acquisition of as much as Elman's neural nets.

But it wasn't easy. Just throwing wellformed strings from the language at a neural net didn't do the trick. After hundreds of thousands of cycles of exposure to random sentences, the neural net had not acquired an ability to predict the end of a sentence (Elman's reasonable test for knowing sentence structure). But most interestingly, Elman found that if he presented the language to the net in doses correlated with sentence length, the net would manage to learn the language tolerably well. That is, the first presented sentences were all simple, only two or three words long; the next batch of presented sentences included a percentage of complex sentences, but the average length was just 3.92 words; the next batch consisted half of simple and half of complex sentences and the average length was 4.38 words. This incremental regime was pursued systematically until the input sentences were typically much longer than a simple clause, and often long enough to exemplify dependencies across several clause boundaries.

Getting the neural net to acquire the language is a step forward; but doing it by managing the order of its exposure to relevant data is a step backward. Children do not undergo such a programmed exposure to input language data. Elman overcame this difficulty by the ingenious step of internalizing the effective ordering of data-experience. Corder (1967) makes a useful distinction between 'input' and 'intake', and this is taken up by formalists (e.g. White (1981:247)) and paralleled by the formalist's distinction between the child's 'primary linguistic data' (PLD) and her 'trigger experience' (See, for example, Lightfoot (1989).).

Elman arranged his neural net so that its intake, or trigger experience, consisted first of short substrings of the input data, then somewhat longer substrings, and so on until there was no restriction on the length of intake strings. This was done by simulating the adjustment of what Elman labelled the system's 'working memory' in such a way that at the beginning it was only capable of attending continuously to short windows of experience. No matter how long were the sentences input to the system, at first its effective intake consisted only of short strings, due to its very limited initial 'working memory'. Now returning to the original unorganized programme of external training, but with the net's internal attention window ('working memory') programmed to start small and subsequently to increase gradually toward an adult value, the net managed to acquire the presented language to a satisfactory level, as defined by a statistical measure of its success at predicting the ends of sentences. This is an interesting and remarkable result. It shows how at a certain stage in life it is actually a disadvantage to have a full adult value for some variable in cognition. (In Section 4, it will be shown that what Elman labels 'working memory' is almost certainly **not** the mechanism that psychologists have come standardly to call 'working memory'; for this reason, when dealing with Elman's construct in this paper, I will continue to use the scare quotes around 'working memory'.)

Extrapolating from artificial neural nets to humans, unless you start with a small 'working memory', you won't start to learn language. But if your 'working memory' doesn't grow to an adult level, you won't finish learning language, either. In a human interpretation of Elman's scenario, a child's 'working memory' expands from very small to an adult value over the period during which language is acquired. (A question that needs to be explored is that of the coordination between the timing of exposure to 'enough' language and this incremental growth of 'working memory'.) In any child, 'working memory' would start small and grow; but if exposure to language only starts when the 'working memory' has expanded significantly beyond its starting value, as may, tragically, have happened with Genie, the child will not be able to acquire the presented language completely. The growth of 'working memory' from infancy to adulthood produces a critical period, after which language acquisition is impaired. Although Elman does not mention the life history timing of the critical period. his work may in fact be taken to predict a very early cutoff point for language acquisition capacity, immediately after trhe first programmed increase in 'working memory'.

3 Marrying the two explanations.

3.1 Evolution of ‘working memory’ growth.

Elman’s account says nothing about **when** in life one should expect the critical period to end; specifically there is no element in Elman’s account which can relate to the coincidence between puberty and the end of the critical period. Elman’s is not an evolutionary account, and so the onset of reproductive capacity at puberty, and its association with fitness, has no place in his model. But Elman’s account is not incompatible with these central ideas in Hurford’s explanation.

On the other hand, Elman’s account of the critical period specifically associates it with a factor in language performance, ‘working memory’, whereas Hurford’s account more vaguely refers merely to a somehow quantifiable language acquisition capacity that could vary according to stage in life. ‘Working memory’ size has the virtue of relative specificity. Evolutionary models, such as Hurford’s, generally make the assumption, which I will not defend in detail here, that linguistic prowess somehow correlates with evolutionarily relevant fitness. This ‘somehow’ needs fleshing out. Replacing ‘amount of language known’, the factor associated with fitness in Hurford’s model, by a factor known to be correlated with sentence-processing ability, would be a step in the right direction.

In Hurford’s account, the evolutionary mechanism giving rise to the decline in language acquisition capacity after puberty is random mutational attrition in the absence of any strong selective pressure to maintain a capacity to acquire something which has already been acquired. This, too, is a weakness, and replacing language acquisition capacity in the model by ‘working memory’ size offers an improvement. ‘Working memory’ needs to grow during language acquisition, but it also needs to stay big during adulthood, so that a person can process complex sentences. If a person manages to acquire language, but then suffers from a decrease in ‘working memory’, his linguistic knowledge will remain intact, but his ability to use it will be impaired. Consequently, there would be evolutionary selectional pressure to retain a reasonably large 'working memory' in adulthood. The variable which, in Elman's model, changes in early life ('working memory') is also the variable which, in an evolutionary account, would be subject to selectional pressure to stay at a high level during later life. An evolutionary account using 'working memory' size as the variable whose life-history timing evolves in the species would in this respect be more plausible than Hurford's 1991 model.

One can envisage, then, a combination of Hurford's and Elman's accounts which answers both the 'when?' question and the 'what' question about the critical period. When? At puberty. What? Increase in 'working memory'. The coherence of this account can be tested by simulations essentially similar in structure to Hurford's 1991 simulations, but using 'working memory' size at various lifestages as the genetically controlled variable correlated with fitness.

Work on such simulations is now in progress (Kirby and Hurford, forthcoming). This work throws into the simulated gene pool factors yielding as phenotypic variants a range of possible settings of 'working memory' size relative to an individual's life history. Assuming a pattern of language acquisition achievement in relation to 'working memory' as discovered by Elman, individuals are exposed to 'the whole language' from the first lifestage onwards, but only acquire knowledge of it (and concomitant ability to process it) as determined by the size of their 'working memory'. Evolutionary fitness is correlated with ability to process sentences – an ability depending on both competence and performance, that is, on how much has been learnt, based on previous 'working memory' profile, and how much can be parsed, based on current 'working memory'.

Among the hypothetical phenotypic variants which compete in this evolutionary scenario, one can conceive of genetically programmed life histories in which, for example, 'working memory' starts large and declines steadily through life, or, as another example, describes a sine curve, rising and falling several times through life (See Figure 4). Neither of these possibilities is likely to survive as the emergent normal phenotype of the species. The expected outcome is a timing of the growth of 'working memory' from very small, reaching maximum value at just the stage in life where the individual becomes reproductively viable, namely at puberty (See Figure 5).

*** FIGURE 4 GOES HERE ***

Figure 4. Two hypothetically possible life-history 'working memory' profiles.

*** FIGURE 5 GOES HERE ***

Figure 5. A life-history 'working memory' profile which would give successful language acquisition within a critical period ending at puberty.

3.2 Implementation: some possibilities

A simulation of the evolutionary factors determining the epigenetic growth of 'working memory', from infancy to puberty, poses some new choices and necessitates some new decisions. These are discussed briefly below, and indications are given of what appear to be interesting questions or conclusions.

3.2.1 Fitness function.

In the earlier simulation, individuals benefitted from simply "having more language". In the new simulation, fitness can be made more specific and more realistic, by adopting Elman's measure of learning success, namely the ability of a learner to process an utterance at a particular point in time, i.e. not simply the internalized competence.

Using this measure of fitness, as opposed to Hurford's (1991) simple 'amount of language acquired', allows us to discriminate correctly between 'working memory' and knowledge of language. Consider two adults who have both learned their language equally well, but one of whom has suffered loss of 'working memory'. While both will have the same competence or linguistic knowledge, they will differ in their performance. The individual with memory loss will be disadvantaged in the evolutionary simulation. In Hurford (1991), experimental runs were carried out in which individuals were highly vulnerable throughout life to stroke-like accidents, simply implemented as substantial decrease in the 'amount of language' possessed by the individual affected. Interestingly, under these conditions, a critical period effect did **not** evolve, a fact that can be explained by the continuing need for individuals, if subject to imminent language loss, for a capacity to re-acquire language at any stage in life (as lizards have the capacity to regrow lost tails).

With a distinction between language loss and 'working memory' loss, more subtle experiments can be carried out in the envisaged simulations, to chart more specifically the kinds of evolutionary circumstances under which a critical period for language acquisition is likely to have evolved. For instance, one may ask, "What happens if individuals are vulnerable throughout life to sudden losses of 'working memory', or of linguistic competence? – Will different epigenetic patterns of 'working memory' growth evolve in these different circumstances?"

It will not be necessary to replicate all of Elman's neural net apparatus. His results provide a function, which we will take as given, which can be envisaged as follows:

$$F(L_i, M_i, D_i) = L_{i+1}$$

where L_i is the learner's linguistic competence at stage i in life (as determined by previous learning), M_i is the learner's genetically determined 'working memory' setting at that stage in life, D_i is the input data, and L_{i+1} is the next state of linguistic competence attained.

An individual's fitness at a particular lifestage will then be a function of his linguistic competence and his 'working memory' size at that lifestage.

'Working memory' is used in language processing. But conceivably the same 'working memory' component of the brain also serves some non-linguistic processing (e.g. spatial work, such as locating the door handle in a room where you've just switched off the light). To the extent that the 'working memory' apparatus that is involved in language processing plays a part in other processing, any fitness function based on 'working memory' should be partly independent of linguistic competence. A small preliminary trial suggests that with some selectional pressure on 'working memory' size from non-linguistic considerations, a more pronounced critical period for language acquisition may evolve. If this turns out indeed to be the case, the reason might well be that selectional pressure for ability to perform non-linguistic tasks pushes for big 'working memory', and this results in an acceleration of the incremental 'working memory' growth due to pressure from language-acquisition. This possibility will be tested more thoroughly.

3.2.2 What is coded in the genome?

The size of 'working memory' at different lifestages may be under genetic control in a variety of ways.

Coding for absolute values. Conceivably, the genome could code for absolute values of 'working memory' at every lifestage, without reference to the value in a preceding lifestage. This seems unlikely, for two reasons. One reason is that this could give rise to abrupt shifts in 'working memory' size, which are unobserved. Another reason is that evolution would have had to take a much more chance-ridden course to achieve a steady increment in 'working memory' size, starting very low in infancy. This may be roughly compared to the chances of throwing a die six times in succession and getting a smooth 1-2-3-4-5-6 sequence of outcomes.

Coding for relative values. It is more plausible that the genes dictate increases or decreases to existing 'working memory' size. That is, for a given lifestage, there could be alleles coding for increments or decrements in 'working memory' size. This would account for the lack of sudden shifts, and would provide a far less spiky fitness landscape for evolution to climb to arrive at a steady growth pattern from infancy to puberty. Nevertheless the space of possibilities would still be large enough for it to be clear that, without natural selection, many other patterns could arise.

Sensitivity to input. An interesting possibility is that, while size of 'working memory' is under some genetic control, it is also sensitive to exposure to language experience. For example, conceivably the steady growth of 'working memory' could be somewhat delayed in the absence of linguistic input; or it might be accelerated in response to abnormally copious amounts of input. The most radical possibility along these lines would be for 'working memory' growth not to be automatic with each lifestage, but instead to respond only to linguistic input. Given this possibility, the organism's 'working memory' would remain immature in the absence of input, and only begin to grow when stimulated by input. Such a proposal would be consistent with recent 'constructivist' ideas; Quartz and Sejnowski (forthcoming) suggest that the growth of dendritic arbors can be promoted by input of the right kind.

Even if size of 'working memory' were determined by response to input in the way just suggested, the mechanism linking 'working memory' size to input would still, of course, be coded in the genes. Evolution has had to choose between an input-sensitive regime for expanding 'working memory' and an automatic regime, where the growth of 'working memory' marches on regardless of how much linguistic data has been experienced. Compromise solutions are also possible, in which there is some degree of sensitivity to input, built into an otherwise automatic program for 'working memory' growth. Whatever the solution, it is coded in the genes, just as it is coded in our genes to reach puberty at a certain 'target' age, but also to overshoot or undershoot this target age given certain environmental conditions, such as malnutrition. Our initial investigations suggest that at least some degree of responsiveness to input may be selected for, so that the evolved program for 'working memory' growth is not entirely automatic and absolutely correlated with lifestage, regardless of the environmental input.

3.2.3 Timing of exposure to language.

Since we are concerned with the 'when?' question, that is the question of at which lifestage 'working memory' reaches its adult value, and the possibility of late language acquisition is therefore effectively denied (as, apparently, with Genie), a decision has to made about how much language is presented to the simulated individuals at each lifestage. In principle, a situation could possibly have evolved in which 'working memory' growth happens very rapidly, say in the first two years of life, so that if enough data were presented, a child could race through the whole process of language acquisition well before puberty. It seems as if rate of exposure to language is another variable to be manipulated in the envisaged simulations. This adds a cultural, as opposed to a biological, dimension to the problem.

4 Empirical data.

Elman gave the label 'working memory' to the recurrent component in his net whose feedback into the rest of the net he experimentally manipulated. In this paper, to avoid confusion with what psychologists have labelled 'working memory', we have used scare quotes around the term for Elman's construct. In tune with Elman's line of research, we accept that it is likely that children are able to perform the feat of extracting a competence grammar for a complex set of sentences by starting with some part of their innate mental machinery set to a low value and programmed to increase over the period to puberty. The present paper proposes an evolutionary explanation for the timing of this maturational increase. But **is** the mental component which thus increases working memory?

Psychologists have found working memory to be a puzzling and elusive phenomenon. There is general agreement that the basic operational test which reveals working memory deficits is getting subjects to repeat an arbitrary series of digits or alphabetic letters. People who have abnormal difficulty in performing such tasks are diagnosed as suffering from a working memory deficit, or 'defective digit span'.

A linguist might be tempted to believe that in processing an utterance, whether as a child acquiring language or as an adult listener parsing a sentence, the mind uses the same temporary storage buffer in which to assemble the incoming words, and that this storage buffer corresponds to working memory. However, a reading of psycholinguistic literature shows that working memory is quite differently involved in these two types of processing. This differentiation weakens the appeal of Elman's model, in which a single component of his system, which he labels 'working memory' plays quite parallel roles in both acquisition and adult processing. Indeed, his model of acquisition is **just** a model of the acquisition of adult parsing expertise. The 'immature' neural net tries to parse sentences and adjusts its weights towards better performance in response to its failures.

When young children are first confronted with language, they do not know (any, or many, of) the words. They receive sequences of essentially nonsense sounds, to which they have to try to assign both structure and meaning (i.e. learn the syntax and semantics of their language). Elman's simulation only replicates the acquisition of (a network representation of) syntactic structure. But, even discounting the lack of semantics in his model, the elements of the input strings (the words) in Elman's simulations are, from the start, less like nonsense words to his system than is the input to a child. Putting aside the nontrivial fact that a child has to locate the word boundaries in the input, the child has to master both the phonological structure of the individual words in the input and the syntactic properties of these words. Elman's neural nets were spared the phonological side of the acquisition task. "Each word was encoded as a vector of 0's in which a single bit was randomly set to 1." (76) In some sense, Elman's nets started with a given vocabulary of words, and what they had to figure out was the syntax (and the insight his work gives us on this task is indeed illuminating).

It seems, from the psychological literature, that what psychologists call working memory is much more clearly implicated in the phonological acquisition task than in the syntactic parsing task. Baddeley (1992) summarizes a number of studies supporting the hypothesis that a 'phonological loop' subcomponent of working memory is centrally involved in the acquisition of vocabulary. Some results that he cites are:

> Baddeley, Papagno and Vallar (1988) ... demonstrated that patient PV, with a very pure short-term phonological memory deficit, showed normal paired associate learning for pairs of meaningful words, together with a severely impaired capacity to learn the

> novel words, needed to acquire items of Russian vocabulary. (284-285)

> (Papagno, Valentine and Baddeley, 1991) and ... (Papagno and Vallar, 1992) [give] results that reinforce the conclusion that the phonological loop [of working memory] is particularly important for the acquisition of novel vocabulary. (285)

> This possibility was explored by Gathercole and Baddeley (1990) in a sample of children who had been selected as having a specific language disability, which involved a combination of normal or above-average nonverbal intelligence, coupled with a delay of at least 2 years in language development. The children did indeed prove to have a particularly marked impairment in the capacity to repeat back material, whether assessed by conventional memory span measures, or in terms of their capacity for repeating back nonwords, varying in length. (285)

> Service (1989) showed in a study of Finnish children that their capacity for learning English was better predicted by their capacity for nonword repetition than by any of a range of other cognitive measures. (285)

One can investigate the degree to which digit span is involved in other tasks, such as parsing complex sentences, by observing the performance of people with a deficit, or by experimentally loading digit span in normal subjects while getting them to perform such other tasks. The results do not implicate working memory in the adult parsing task as clearly as one might expect.

> There is some evidence to suggest that it [a component of working memory] plays a role in speech comprehension, although most STM deficit patients are impaired on processing only relatively complex sentences (Vallar and Shallice, 1990). ... Typically, although such subjects have a digit span of only one or two items, their span for structured sentential material tends to be six or seven words. (Baddeley, 1992:284)

Caplan (1987) also poses the question of "what memory system(s) is (are) used in sentence comprehension" (319), and reviews three case studies, leading to the conclusion:

> We thus see that the study of syntactic comprehension disorders in patients with expressive agrammatism and auditory-verbal STM impairments, which originally was taken to show that these supposedly "primary" impairments cause "secondary" disturbances in assigning and comprehending syntactic form, in fact suggests that syntactic comprehension impairments are often independent primary disorders of sentence processing. (323)

What (psycho)linguists interested in the parsing process have called working memory involves more than just a place for temporary storage of short unstructured strings (such as digits). It is also a place where partially analyzed material is stored. "There is good evidence that a sentence is divided into constituents in working memory" (Clark and Clark, 1977:52). Clark and Clark cite the classic probe-latency work of Ammon (1968) as such evidence. Wanner and Maratsos (1978) use the term 'transient memory' for the store used in sentence comprehension and test whether a task involving remembering lists of names (a classic psychologist's working memory task) interferes with this transient memory; it does to some extent, and more at some points in sentences than others. Wanner and Maratsos suggest a relationship, but not an identity, between sentence comprehension and memory for lists of names. In Elman's recurrent net, the extent of 'working memory' was the amount of coded and 'digested' information about previous elements in a string, stored in a 'context' layer. This seems closer to the traditional psycholinguistic (e.g. Clark and Clark) idea of working memory, for which, for most of the previous discussion, I have used scare quotes.

The uncertainty over the degree to which working memory (as defined by performance on digit span tasks) is involved in adult sentence processing may possibly vitiate one component of the evolutionary account proposed in this paper. This account postulated adult sentence processing as a selective pressure to keep 'working memory' at a high value during adulthood. If it turns out that 'working memory' is not, or little, involved in sentence processing, pressure from a different function will have to be sought, and not necessarily a specifically linguistic function (recall the final paragraph of subsection 3.2.1 above).

5 Conclusion

Where does this leave us? Clearly, the mechanism in his simulations that Elman labelled 'working memory' is not the same thing as psychologists have in mind when they write about working memory, although the latter is implicated in an interesting way in language acquisition. It seems reasonable to believe that Elman's results with artificial systems tell us something about the way natural organisms work. It is also likely that natural systems are very much more complex. We can accept the very general message of Elman's work, about "the advantages of starting small", a message which is echoed from many quarters in developmental psychology.

Bjorklund and Green (1992) give a concise review of work on "the adaptive nature of cognitive immaturity" across four fields; metacognition ("a person's knowledge about his or her own cognitions and the factors that influence thinking" (47)); egocentricity; plasticity and speed of cognitive processing; and language acquisition. The work in language acquisition most closely associated with the advantages of starting small, or otherwise incomplete by comparison with adults, is Elissa Newport's. In a series of studies, she and colleagues have shown the plausible learning advantages of limited resources in children (Newport, 1990, 1988; Goldowsky and Newport, 1993).

Whatever it is in the infant that 'starts small' and then grows, its growth facilitating language acquisition, the rate of this growth seems likely to be controlled in large part by genetic factors. And the evolution of the genotype that codes for this growing phenotype is likely to be the product of Darwinian natural selection. This paper has shown how the evolution of such a growing trait (here labelled ‘working memory’) can be modelled, with an adult value being reached around puberty. When psycholinguists get closer to discovering the actual mechanisms in the child that ‘start small’, facilitating language acquisition, there is no reason to believe that a similar evolutionary account will not be feasible. Such evolutionary explanations help to fill in the whole picture, not just of how humans work, but of how they got to be that way[1].

[1] I thank Simon Kirby for good advice on this paper, and David Pesetsky for asking a question which significantly changed its shape; responsibility for it is mine, however.

6 REFERENCES

Ammon, P.R. 1968. "The perception of grammatical relations in sentences: a methodological exploration" *Journal of Verbal Learning and Verbal Behavior*, 7: 869-875.

Baddeley, Alan. 1992. "Working memory: the interface between memory and cognition". *Journal of Cognitive Neuroscience*, 4,3: 281-288.

Baddeley, Alan, Papagno, C., and Vallar, G. 1988. "When long-tern learning depends on short-term storage". *Journal of Memory and Language*, 27: 586-595.

Bates, Elizabeth. 1994. "Modularity, domain specificity and the development of language". *Discussions in Neuroscience*, X, 1/2:136-149.

Bates, Elizabeth, Thal, Donna, and Marchman, Virginia. 1991. "Symbols and syntax: a Darwinian approach to language development". In: Krasnegor, Norman A., Rumbaugh, Duane M., Schiefelbusch, Richard R., and Studdert-Kennedy, Michael (eds), *Biological and Behavioral Determinants of Language Development*, 29-65. Hillsdale, NJ : Lawrence Erlbaum.

Bates, Elizabeth, and MacWhinney, Brian. 1990. "Welcome to functionalism". *Behavioral and Brain Sciences*, 13,4: 727-728.

Bjorklund, David F., and Green, Brandi L. 1992. "The adaptive nature of cognitive immaturity". *American Psychologist*, 47,1: 46-54.

Caplan, David, 1987. *Neurolinguistics and Linguistic Aphasiology: an introduction.* Cambridge: Cambridge University Press.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge, MA : MIT Press.

Chomsky, Noam. 1968. *Language and Mind.* Harcourt, Brace and World.

Chomsky, Noam. 1980. *Rules and Representations.* Columbia University Press.

Chomsky, Noam. 1982. *The Generative Enterprise: a discussion with Riny Huybregts and Henk van Riemsdijk.* Dordrecht, Holland: Foris Publications.

Chomsky, Noam. 1988. *Language and Problems of Knowledge: the Managua lectures.* Cambridge, MA : MIT Press.

Chomsky, Noam., and Lasnik, Howard. 1977. "Filters and control". *Linguistic Inquiry*, 8,3: 425-504.

Clark, Herbert, and Clark, Eve, 1977. *Psychology and Language: an introduiction to Psycholinguistics.* New York: Harcourt Brace Jovanovich.

Corder, S. 1967. "The significance of learners' errors". *International Review of Applied Linguistics*, 5: 161-70.

Elman, Jeffrey L. 1993. "Learning and development in neural networks: the importance of starting small". *Cognition*, 48: 71-99.

Friedman, Joyce, 1968 *A Computer System for Transformational Grammar*, CS-84, AF-21. Computer Science Department, Stanford University.

Gathercole, S. and Baddeley, A. 1990. "Phonological memory deficits in language-disordered children: Is there a causal connection?". *Journal of Memory and Language*, 29: 336-360.

Goldowsky, B.N., and Newport, Elissa I. 1993. "Modeling the effects of processing limitations on the acquisition of morphology: the less is more hypothesis". In Clark, E. (ed.) *The proceedings of the 24th Annual Child language Research Forum* 124-138. Stanford, CA: Center for the Study of Language and Information.

Hawkins, John A. 1990. "A parsing theory of word order universals". *Linguistic Inquiry.* 21: 223-261.

Hurford, James R. 1991. "The evolution of the critical period for language acquisition". *Cognition*, 40: 159-201.

Kirby, Simon. 1994. "Adaptive explanations for language universals: a model of Hawkins' performance theory". *Sprachtypologie und Universalienforschung.* 47: 186-210.

Kirby, Simon, and Hurford, James R. Forthcoming. "Evolution of incremental learning: working memory and the critical period" (provisional title).

Larson, Richard K., Warren, David S., Freire de Lima e Silva, Juliana, and Sagonas, Konstantinos. 1996. *Syntactica* (NeXTStep Edition) Cambridge, MA : MIT Press.

Lightfoot, David. 1989. "The child's trigger experience: degree-0 learnability". *Behavioral and Brain Sciences*, 12: 321-334

Newport, Elissa L. 1988. "Constraints on learning and their role in language acquisition: studies in the acquisition of American Sign Language". *Language Sciences*, 10: 147-172.

Newport, Elissa L. 1990. "Maturational constraints on language learning". *Cognitive Science*, 14: 11-28.

Papagno, C., Valentine, T., and Baddeley, A. 1991. "Phonological short-term memory and foreign language vocabulary learning" *Journal of Memory and Language*, 30: 331-347.

Papagno, C., and Vallar, G. 1992. "Phonological short-term memory and the learning of novel words: the effect of phonological similarity and item length". *Quarterly Journal of Experimental Psychology* 44A: 47-67.

Piattelli-Palmarini, M. 1989. “Evolution, selection, and cognition: From ‘learning’ to parameter setting in the study of language”. *Cognition*, 31: 1-44.

Pinker, Steven and Bloom, Paul. 1990. “Natural language and natural selection”. *Behavioral and Brain Sciences*, 13,4: 707-727.

Quartz, Steven R., and Sejnowski, Terrence J. Forthcoming. “The neural basis of cognitive development: a constructivist manifesto”. *Behavioral and Brain Sciences.*

Service, E. 1989. “Phonological coding in working memory and foreign language learning.” University of Helsinki, *General Psychology Monographs*, No.B9.

Wanner, Eric, and Maratsos, Michael. 1978. “An ATN approach to comprehension” In Halle, Morris, Bresnan, Joan, and Miller, George A. (eds) *Linguistic Theory and Psychological Reality* 118-159. Cambridge, MA: MIT Press.

White, L. 1981. “The responsibility of grammatical theory to acquisitional data”. In Hornstein, Norbert, and Lightfoot, David (eds.), *Explanation in Linguistics.* 214-71. London: Longman.