Embedding parameters in a model of learning language. Draft, October, 2017

T. G. Bever

What cannot be denied is that any organism that extrapolates from its experience does so on the basis of principles that are not themselves supplied by its experience. Fodor, 1966

….using parameters to characterize the space of possible languages does not in itself provide a model of learning. Pearl and Lidz, 2014

[we] formulate the notion of "learning/c"…. acquiring structures by combining basic linguamental structures (languages of the mind) ….if induction fails in general, it fails to account for the acquisition of language…learning/c…which recruits innate structures….[so] the claim that language is innate, depends on….innate structures that cannot be accounted for by the mechanisms of learning/c, [not induction] Bever, Carroll and Miller, 1976,

Introduction and Summary

An enduring question raised by Eric Lenneberg's papers and ultimate book on the biology of language, is the extent to which the mechanisms involved in linguistic knowledge and behavior are typical of biologically based species specific processes. In the subsequent fifty years, the particular architecture of what language is has evolved in half a dozen stages, with increasing simplicity being the general trend. Models of its acquisition, however, have varied, one of the most elaborate being "principles and parameters". In that model, a small set of specific principles are innate; but the real work of specifying the innate basis for acquisition rests on a large number of binary choices for particular languages – so-called parameters.. "Learning" is the result of interaction of those dimensions with language experience that "sets" each parameter according to the particular language. In contrast with this, the "principles" and basic architecture of language are not learned from experience, though they may depend on general experience for them to have any effects at all.

That is, the architecture of language and its principles are not learned, but innate, while specifics of each language are automatically learned by setting parameters.

Parameters have been proposed because they solve various related problems:

- how are specifics of individual languages learned so fast?
- how can specifics of language be learned with so little data?
- why do languages differ typologically in characteristic, but limited ways?

In this paper I address a mystery hitherto ignored, namely what is the language learning child actually doing dynamically, that results in appearing to set parameters. I outline how a general problem solving model of learning can account for how the child integrates experience with knowledge: insofar as parameters are not themselves learned. The dynamic model itself rests on what seems to be a unique property of humans – the desire to interpret the world as a problem to be solved.

The result is that what remains intact is the concept that the architecture and processes of languages are innate in the sense that Lenneberg intended: they emerge to a great extent maturationally, given minimal general experience with language. The particulars of a given language set the problem that the language learning child solves because using available mental architectures and processes to solve such problems is a basic drive of human cognition.

Previous discussions of Parameters

This paper was stimulated by articles on parameter setting to be included in a special issue of this journal, edited by Simin Karimi and Massimo Piattelli-Palmarini. Most of the papers were pro-parameters, as descriptive devices and as innate structures that provide an acquisition model that accounts for its rapidity and success, given The Poverty of the Stimulus. A few articles are critical of parameters as descriptive and at least one, by David Lightfoot proposes an alternate acquisition model. I take Lightfoot's proposal (as I understand it) to be the following.

a) The child has an I-language "toolkit", with which s/he can construct a (derivational?) grammar.

b) The child has an available (innate?) parser which at the least creates constituency out of what s/he hears. This part of the model seems consistent with a form of induction.

c) The child then (intuitively) uses the toolkit to organize consistencies in the output of the parser into an instance of I-language, specific to the inputs.

d) (corollary) The child adapts to changes in the language, incorporating/creating novel I-languages over time.

Lightfoot uses this model primarily to address how languages can vary and change rather than being imprisoned by particular parameter setting accidents. If this is a correct summary of his views on acquisition, I find that it resonates with my erratic attempts over many years to put together a coherent theory of acquisition, with a reduced role for parameter setting, which always struck me as too abstract at best, and theoretically problematic at worst.

It is arguable that parameters within a given grammatical framework provide a meaningful basis for differentiating language typologies. Given a small enough number of parameters, the typological framework can be interpreted as a potential constraint on possible languages; this in turn can be taken as knowledge that children bring to learning their first language. The goal of this paper is to provide a dynamic learning model that relegates the role of parameters to formulating possible grammars as potential solutions to problems set by the surface forms that children encounter. That is, on this model, the child's proximal problem is to access its derivational formulating ability to reconcile apparent contrasts (contradictions) in

patterns of input it recognizes: "settings" of parameters are the residue of successful derivational solutions of the surface problems.

I have tried to elaborate version of an acquisition model that answers a number of theoretical and empirical questions, in many cases not directly addressed by the usual parameter setting acquisition models. At the outset, I leave open whether parameters are uniquely innate, follow from general cognitive structures, or from universal extrinsic computational laws. I accept the possibility that they can helpfully define dimensions that can play a role in grammar construction. What I question is their direct causal role in the actual dynamics of language learning: I try to show that those dynamics depend on a distinct motivational process that explains a range of facts about language and its acquisition. How or whether parameters are themselves, or whether their definition is a function of the learning process itself, is a further empirical question.

1. Aside from language structure itself, ***what potentially innate cognitive capacities differentiate humans from other animals***?

I (among others) have suggested that **humans naturally and implicitly organize experiences as problems to solve (e.g.,** Wertheimer). This shows up in implicitly preferred activities, e.g., in the infant's simple game of peek-a-boo, and in adults' implicit preferences for particular objects, e.g., the golden mean rectangle. In explicit aesthetic objects it can show up e.g., in classic dramas, or in music. In each case, analysis shows that there is a representational conflict that is resolved.

A classic problem solving model related to this process derives from gestalt psychologists: an emerging representational conflict between 2 aspects of a situation is resolved by accessing a different dimension of analysis. The paradigmatic example is from Wertheimer. Subjects (in the 1930s) found how to use x-rays to kill tumors to be a conundrum: too strong a beam would kill intervening tissue, too weak a beam would not kill the tumor. The solution was to think of the gun as moving in a circle around the patient, with the only constant being the focus on the tumor location, that is, to reinterpret the use of the x-ray in a second (or third) dimension.

Most important to my consideration here is the invocation of the "aha" reaction that people have when they think they have solved a problem. **Perhaps what is innate is this pleasurable mechanism**, explaining why humans like to create and solve problems. (Evidence for general problem solving in animals is scant, doing it for its own sake is unattested – but might be difficult for us humans to recognize)

2**.** ***Is it necessary to assume a special learning mechanism for language, or could it be an adaptation of a more general cognitive system such as problem solving***?

My proposal is that **acquisition can be viewed as dependent on an unconscious application of conscious problem solving in general**. On this model, the child accumulates statistical generalizations based on simple parsing structures (e.g. lexical categorization, adjunction, location of phrasal heads); the most notorious

example is nvx=”agent predicate other”. When the child experiences a violation of a generalization, e.g., “harry got hit”, this “problem" triggers access to an independently available I-language “tool kit” which can differentiate constructions by assigning different derivations to them: this involves accessing a different level of representation, which is a processes involved in aha-inducing problem solving. So, on this model, the tools to construct I-languages are available/innate, and are accessed as stimulated by induced representational conflicts in E-language. This fits the problem solving model perfectly.

It is possible that at least some major parameters are part of the I-language toolkit, setting constraints on possible derivations. What is at issue here is that parameters do not directly explain the dynamics of acquisition, even if they do play a role in formulating derivational hypotheses.

3. ***What is the motivation for the child to learn the complex abstract structures of an I-language***?

There have been two standard answers to this question. a) the child wants to communicate with fellow humans; b) I-language grows without extrinsic motivation, it is simply what maturing kids do, like learning to walk. One view is based on external pressure, the other on intrinsic automatic mechanisms. I am proposing an intermediate view. Namely, **humans intuitively seek problems to solve, motivated by an internal urge to experience even small implicit “aha” reactions**.

This may couple with a more pervasive mechanism in the animal world, children want to be like the adults (remember Phil Harris singing, “I want to walk like you ….I want to talk like you” in the film Mowglis). So the problem that language sets, is how to be like adults. The mechansim for solving it bit by bit is applying the internalized problem solving strategy, which in turn elicits mini-ahas, which are motivating because we are problem solving and aha-loving animals. So learning/constructing I-language is one micro-fun-event after another.

4. ***What prior parameter-free models of the acquisition process presage the model?*** Many models of acquisition relying on induction alone, finesse the problem of acquiring a deep structure of some kind, and derivations relating it to surface forms, by simply denying that anything like I-language exists (E.g., Bates and Macwhinney, many connectionist models, classic SR models, even mediation models (Osgood), and so on). An early exception to this is Fodor’s 1966 essay on how language deep structures might be learned starting with inductive procedures to build up surface representations. He suggests that some form of induction can provide repre-sentations of surface forms that are similar but can only be related via deep structure derivations. For example if a child encounters a set of cases like (i) and (ii) along with information that they roughly cover the same situation, this can become a surface contrast to resolve by assuming a common derivation. Similarly, sets of cases like (iii) and (iv) demand a derivational solution even more strongly because of the notable rearrangement of the word order and associated verb agreement processes.

(i) Mommy called daddy up

(ii) Mommy called up daddy

(iii) Mommy found Daddy's car keys
(iv) Daddy's car keys got/were found (by Mommy)

Fodor leaves open whether this kind of learning is unique to language or is an application of a more general learning algorithm. Here I suggest that it is indeed a specific application of general problem solving procedures and motivation: invoking an inner form of sentences to relate their outer forms is an example of problem solving a la Wertheimer. Today we have a vastly richer notion of I-language and a small 'minimal' set of operations that it specifies. One contribution in the discussion here is suggesting that the problem solving model motivates finding derivations for distinct surface forms that are thematically related.

5. ***What kinds of facts does this model potentially explain?***

A. First a set of facts, none of which are actually explained by I-language constraints.

a) **Attested languages are a narrow subset of I-language possible ones**. This follows from any model that requires that the language be parsable by a child, that it have inducible E patterns and memorable I-structures that are easily constructed. Parameter setting models have a difficulty that in fact they generate an extremely large set of grammars, without any explanation of which parameters or sets of parameter settings are blocked (the one clear attempt at formulating this is Baker's hierarchical architectural routing of parameter decisions, but see MacCauley, "30 million theories of grammar").

b) **Languages have canonical forms**, at every level of representation, phonemic, morphological, and syntactic. This seems to be a universal property, wherever one has looked. Of course, each language has its own canonical forms, e.g., the dominant syllabic type, dominant morphological affixing patterns, dominant syntactic patterns (Slobin and Bever). I-Language does not specify any of these, indeed does not require that canonical forms exist at all. This elicits a particularly puzzling aspect of the facts: language would be much more efficient on information theoretic grounds if each kind of structure were equally frequent with every other, rather than having a strongly dominant form.

But the acquisition model requires that at least some patterns be susceptible to induction, since that is a critical component of the problem solving process. Without that, the acquisition model cannot get off the ground.

The canonical syntactic structure has a surface and a thematic structure.
In English, the surface pattern (ignoring interjections adjuncts, etc.) is:

v) Noun-i ,Verb (agreeing with Noun), other stuff (NP, PP, adverb, complement)

vi) The thematic mapping pattern is: NVN = Agent/Experiencer, Predicate object (of some kind)

The surface pattern covers a significant range of. Importantly**, not all of those constructions conform to the thematic mapping pattern** (e.g., passives, object relatives)**.** Thus, the outer structure of the canonical form is multiply represented, serving initial inductive processes: as the child starts to experience violations of the thematic mapping, this stimulates the representational conflict – a problem to be solved by accessing the I-language tool kit to derive the differences.

c) **Languages tend to have some patterning "harmony**". This is a corollary of (b), though not ineluctable (hence the use of "some" patterning). If a vp/sentence is head final, it is frequently the case that phrases with other categories, e.g. NPs, are also head final. If parsing adopts strategies that transcend particular phrase categories, "head" vs "complements" would be a pervasive pattern (note this is what one might expect if the parsing strategies themselves are based on general principles). So an early stage of establishing constituency would be facilitated if Head+X. or X+Head is pervasive (Hawkins' EIC). Hoswever, this is not a structural principle, but a behavioral one, hence subject to variability, and not a parameter (as several articles in the special issue point out). (See Newport and 2015 for behavioral evidence that harmonic parallels appear strongly in normal acquisition by children, more so than adults learning a second language).

d) **Certain commonly attested language features that seem to be structural but must be stipulated in UG may be explained** by the constraining power of the canonical form. An English (and other similar languages) phenomenon is the EPP (extended projection principle), in general requiring that surface subject positions be filled. As noted, the canonical form in English has two levels of analysis not just a thematic mapping structure: it has a surface pattern (v) which requires a surface subject. On this interpretation, EPP may well be a structural property of English and other languages, but not one stipulated in UG.

However, some arguments consistent with this have been couched within a UG framework. For example, it has been suggested that the EPP is a constraint on the Phonetic Externalization Interface, not on the merge-formed syntactic structures themselves. (Holmberg, Landau, Abe). More recent attempts to embed EPP as a natural consequence of syntactic principles have been suggested by Rizzi and Chomsky. Their proposals link the EPP to the ECP via feature manipulations in particular structural configurations. This involves a kind of cartographic structure, and it remains to be seen if that is itself just a different kind of stipulation within UG.

e) Young children show a general pattern of acquiring statistically valid generalizations in many cognitive domains, including language. For example, two year olds perform better on standard Piagetian conservation tasks (eg liquid poured into a low fat glass vs. a tall thin one) than 3-4 year olds: the argument has been that

the older kids have built up a generalization that if something looks longer or bigger it actually has more "stuff" (Mehler and Bever); similarly, in language two year olds understand object first clefts and passives better than chance and much better than 3-4 year olds: the proposal for that is that the older children have acquired the generalization that the first encountered noun is the agent of the predicate (Bever, Mehler. Valian). Such phenomena (now replicated in different ways) provide empirical support for the inductive component of language acquisition.

B. ***Second. a set of facts about language behavior***.

a) Strategies and canonical forms **help with the POS, Plato's problem**. The number of complete grammatical sentences a child hears is relatively small, perhaps 10,000/year at most if some statistics about caretakerese are to be believed (and this does not start until two years, younger caretakerese tends to have mostly a lot of googoo talk in it). Furthermore, the number of sentences does not seem to correlate with how quickly a child advances in language, at least at early stages (Newport et al, C.Chomsky). If the child builds up templates this can enable the child to create its own well-formed sequences without yet having an I-grammatical structure for them. This may be the explanation of the discovery that children talk to themselves, apparently practicing different versions of a given sentence (e.g., Ruth Weir's "Language in the crib"). In this way, the child can use its induced templates to generate sentences that it in fact never heard, greatly expanding the data set it can consider. This is rather what like linguists do when formulating grammars, and comports well with e.g. Valian's notion of the child as "little linguist".

b) **Caretakerese tends to be short sentences**, with a small number of types (English examples), yes/no questions, imperatives, statements. Short sentences have the virtue that they characteristically explicitly display all or part of the canonical form, while longer complex constructions can obscure it. So, the child receives enough dominant exposure to the canonical form, whether the caretaker is talkative or not. Short sentences, however, characteristically do not provide evidence for constraints that span clauses/ propositions.

c) The inductive system **can utilize incomplete sentences, false starts or simple phrase fragments**, especially if they display parts of the canonical form. False starts may be particularly helpful since they tend to be isolated phrases, giving useful information about surface patterns, free of thematic mappings (Levelt). So, data relevant for acquisition does not have to be fully formed grammatical sentences and some ungrammagtical sequences may be felicitous for the inductive system.

   The aspect of caretakerese that is critical may be its exaggerated intonation patterns. This will offer an attentive child information about likely surface constituents, that are also likely to map onto thematic conceptual structures.

If a caretaker offered the child a segmentation as in (i), this would not be helpful, but as in (ii) immediately conveys what surface sequence to map onto a conceptual representation. I'm not suggesting that this feature of caretakerese has not been noted before (e.g., Kuhl et al); rather I am noting how it fits into the problem solving model.

(i) (The farmer's in)(the dell Pussy's in) (the well)
(ii) (The farmer's) (in the dell) (Pussy's in the well)

C. ***The model offers an integrated dynamic theory of language change***

An early example of this is our paper on the historical loss in English, of the option of deleting subject pronouns/complementizers (paper attached). No doubt, we over-simplified, or even misrepresented the facts. But the basic framework to explain the change was the following.

a) The child builds up processing strategies based on frequent constructions.
b) The strategies constrain the child's grammatical hypotheses.
c) Old English was richly inflected, both marking noun case and verb agreement, with many classes of both nouns and verbs, each with its own set of suffixes.
d) Following (c), marking the beginning of a subject relative clause with a complementizer was optional. Thus, (i) and (ii) were both ok, in those cases where the verb agreement case marking uniquely blocked misinterpreting the first noun as the subject of the relative clause verb.

i) The boy (that saw you) is outside.
ii) The boy (saw you) is outside.

e) In the 11th-12th centuries Old English rapidly lost most of its inflections. (The frequently assumed cause was the massive influx of Fench words, unmarked as to their lexical class, treated as the neutral case - this created a lopsided lexical array with one overwhelming simple class and a fractionated set of other classes (referred to retrospectively as "strong" in the case of verbs. Memory constraints set in, so the older inflectional system collapsed for nouns, and mostly for verbs.)
f) Without inflections, word order became increasingly characteristic, moving towards the modern nvx = agent predicate other….
g) The child induced a corresponding strategy to (f).
h) Sentences like (ii) became difficult to parse compared with the still available (i)
i) The child adapted, creating a grammar which disallowed (ii) but still allowed no complementizer in sentences like iii and iv: they did not induce local ambiguity.

iii) The boy who you saw….
iv) The boy you saw….

As I said, no doubt we oversimplified or misrepresented the actual facts. But the theory seems consistent with Lightfoot's view on how the child adapts to language change, creating new grammars with its I-language tool box that conform to

the parsing patterns.

F. Comparing assumptions with Parameter setting models of acquisition.

Have I enriched the set of assumptions about acquisition in relation to parameter setting models? I don't think so. Here I contrast the problem solving model with parameter setting models that purport to have the entire acquisition process depend only on parameters and triggers that set them either with one-time exposure or statistically preponderant exposures.

Both models assume an available parsing routine of some kind that renders at least hierarchical structures paired with thematic relations. Parameter setting models ordinarily require that the parse be labeled correctly with linguistic universal categories (though this may have been more a matter of convenience than necessity). The problem solving model does not require that, only that the structural configuration be reliable and have some kind of labeling, at last based on lexical categories.

Both models assume that lexical items are available in the input with some kind of categorization. (I think that) parameter setting requires that the categories be from the set of universally available syntactic categories. The problem solving model does not: so called semantic categories may be sufficient initially, before the I-language toolkit is called into service (whence the actual syntactic categories, N, V, adj, etc. are to be found).

Both models can predict harmony, although in each case the assumption has to be that the structures in patterns are labeled in such a way that it is possible to express the constraints across different categories.

The parameter setting model says nothing intrinsically about utilizing fragments, ungrammatical sentences, false starts and so on, though some models can stipulate how such input can be used. It is intrinsic to building up patterns in the problem solving model, that partial or semi-grammatical input will be automatically utilized when it overlaps with larger structures.

The parameter setting model makes no particular use of the child's genius at word learning, especially "fast mapping". The problem solving model can take each word as the limiting case of mapping thematic structures onto surface forms: \indeed the fact that most words can appear in different constructions, can constitute part of the problem that language learning sets: the child can resolve the different uses sometimes by accessing the under-lying lexical category – it may be equally important that the child has to accept that some lexical items are truly ambiguous, hence have no resolution. Both models assume that there is some reliable semantic/ thematic representation that relates to the surface form. But instantiations of parameter setting models require that the child have a consistent and accurate thematic representation of each sequence: the problem solving model can accept more

slop and approximation, since the inductive process involves converging on stimulus patterns,

The usual parameter setting model in principle only needs a critical "trigger" in the input, while the problem solving model requires some statistical reliability to build up initial patterns. Note that Sakas, Yang and others have argued that the single "triggers" proposal is neither tractable logically, nor does it fit the facts. They propose in fact, a statistical model on which evidence accumulates to strengthen one or the other setting on each parameter. When the probability of a setting reaches a high criterion (never reaching 100% of course) then "the learning system" treats that as categorical. Thus, their model still requires a categorical setting on each innate parameter as the final arbiter of acquisition, but one that takes input statistically as evidence.

The parameter setting model assumes that the child has some extrinsic motive to learn the I-language and parameter settings, e.g., urge to communicate, or to be like the adults. Alternatively, some models assume that as innate structures, parameters inherently demand of the environment data to set them, and this constitutes an automatic mechanism for a grammar to "grow" in the mind. The problem solving model allows for a range of extrinsic motives: but the critical one is the child's intuitive zeal to treat language as a problem, and to get enjoyment out of figuring it out as a puzzle.

Parameter setting models express how learning works in terms of the interaction of the parameters with input. The problem solving model intrinsically attaches internal computation as an intrinsic *experience* of the language. The distinction between input and experience is critical: input alone is useless, it has to be embedded and organized within the mind to become useful experience. It can be argued, I suppose, that setting a parameter – even turning a statistically preponderant value into a categorical one – is a mental experience. But I find the operation of actually taking the input as a problem to be solved necessarily enfolds it within a set of operations as experiences.

G. ***Big unsolved related activities and problems***:

a) Have we replaced brittle and mysterious innate parameters with **a flexible mysterious innate parser?** In order for this program (like Lightfoot's) to work, there must be parsing principles available independent of any specific language – the last thing we want to do is introduce parsing parameters. Induction over word strings is not adequate for this model – there have to be at least what used to be called, some kind of "Surface Structures" over which induction can operate, that is structure dependence must be an immediate aspect of building up statistically supported patterns. There may be several mitigating factors in favor of a language independent independent parser if there is one.

   i) Most important, **the initial parses do not have to be grammatically**

correct. Rather, they must be "good enough" , and reliable in the sequence structure so induction can range over them.

ii) We know today that **infants are exquisite induction machines**, over even small numbers of instances (Saffran, Gerken, etc.). There is also clear evidence that statistical generalizations are acquired in many cognitive domains during early childhood. (e.g., demonstrated in u-shaped developmental functions).

iii) The parser has **fewer principles and moving parts**: particular parsers independent of particular languages include proposals by Kimball, Frazier, JDFodor, Hawkins?: in each case a few processes are invoked to build labeled surface constituencies, which in our model is the basic input to the inductive component. (Since correct structures are not required, this may be a proper application of x-bar theory).

iv) **Initial parsers may be composed out of general inductive and binding mechanisms**. General "Structure building" hypotheses about how linguistic trees are constructed have been proposed as creating trees by Gernsbacher, Bates&Mcwhinney and others. More recently, "Fluid construction grammar" proposals claim that a small set of general automatic representation forming procedures can produce acceptable tree representations. And of course, many variants of connectionist nets have been invoked as general learning devices that can simulate patterns (Remember again, that the surface level induced patterns do not have to be syntactically correct, only reliable enough to provide a summary of some aspects of a language, and some conflicts with other aspects.)

-

b) The wrts aren't there, and **the POS is a lot worse than we thought.** Charles Hockett famously likened the comprehension and production of sentences as involving a continuous just cooked omelet on a moving belt created by cracking successive eggs above it. Going a bit further than his original metaphor, the perceptual system has to decode the scrambled omelet in which the eggs run into and over each other, and the production system has to break the eggs in such a way that the perceptual system can do that. So the actual acoustic manifestation of underlying discrete units and structures is often seriously degraded. Recent investigations of normal conversation, including caretakerese, suggest that in some cases the "produced" or "perceived" words are totally absent phonetic-acoustically – what is left are traces and cues of the underlying (surface) representations. Most important, in some cases the resolution of those cues does not fall into place until after their presentation, based on later acoustic input. This means that the assumption that the language acquiring child gets complete information about the words and their order is overly generous to every model of how the child's language experience shapes its acquisition.

c) Two levels of problem solving in language, synchronic and evolutionary.
The preceding discussion concerns the acquisition of language in individuals: each time an induced SurfaceForm->ThematicRelations pattern is violated, it

creates a conflict, namely a problem to solve by accessing the capacity to create differentiate derivations that arrive at the same surface form.

A deeper question is where the I-Language toolkit itself comes from, i.e. how it satisfies some internal mental dynamics as well as being shaped by extrinsic forces. To speculate on this we can refer to the basic conflict between the seriality of language output with the hierarchical complexity of the conceptual representations. Various important linguists have noted this duality as setting the boundaries on possible grammars (Wundt, Tesniere, Saussure, Chomsky, even Hockett in his way). Essentially the two kinds of representations are incompatible, "requiring" a grammar as a pathway between them.

The considerations here offer the potential for a dynamic explanation of the presence of the I-language toolkit itself. On this view, it may not be that human cognition uniquely embraces the duality – arguably some animal "communication" systems exhibit it, without any evidencc of a grammar (Pace Berwick et al). Rather, the extreme nature of the duality is sharpened by a separate capacity that appears unique to humans – the ability to recognize the surface/meaning relation as it is expressed in distinct words – in my opinion, too little attention has been given to the potentially extrinsic prodigious capacity to learn individual words. At the least, a particular word is the limiting case, which maps a conceptual structure onto a single surface element as opposed to a string. The duality exists in such cases, and becomes a problem when words interact with each other at the surface level.

On the problem solving view, the role of some kind of grammar in humans, is that human cognition implicitly organizes the duality as a problem to solve. By assumption, humans solve the problem by accessing a different kind of set of processes, which at least provide some stable intermediate grammatical structures, constrained by cognitive capacities, computational efficiency and other interacting factors. How much of this is uniquely innate for language, how much follows from general cognitive capacities, how much from extrinsic constraints (physical laws; computational simplicity) remains for discovery. As always, the purpose of these considerations is to further focus on what we are looking for.

d) **What is the relation between this model and models of phonological acquisition?** Strictly speaking, this essay is about acquiring syntax, but it may have implications for those models of acquiring phonology that assume it is not strictly phonetics-based but involves contrasts at an abstract computational level (classically Jakobson and Halle, Martinet, more recently, Dresher). The historical long term approach to learning a language's sound system has been to use information about contrasts, and lack of contrasts between phonetically similar elements to determine the underlying phonemes as a solution. In the problem solving terms described here, each contrast

can be thought of as a potentially conflicting representation, to be "resolved" by accessing a representation of a feature or phoneme. This process can directly involve phonological derivations underlying the surface phones, demonstrating that the process involves abstract representations (Dresher). (Thanks to Lightfoot for pointing out the notion of contrast in phonology acquisition - this is mostly a promissory note for future investigation).

e) What is the relation between this model and other characteristically human behaviors*?* One way to examine the extent to which the model is unique to language is to examine other characteristically unique human behaviors. Of course, explicit problem creation and solving is an intrinsic component of human social life: unlike animal groups in general, humans invent games, competitions, puzzles, usually with no intrinsic or practical value; rather they activate the fundamental enjoyable capacity to solve problems. Explicit fine arts, music, painting, theatre, dance also involve tension and resolution. But assessing the value of particular art works is contaminated by their commercial role in societies.

To remove the social function of art from consideration, I have examined how natural aesthetic objects of everyday life, elicit mental representations that fit the problem solving model. This follows Fechner's argument that such examples reveal how the mind works when running freely (im Leerlauf) without being constrained by practical applications. Examples I have examined include the song "Happy Birthday", the golden mean rectangle, the tattoo "shave and a haircut", the infant's game of "peek a boo" – the general argument is that such aesthetically stable objects elicit an unconscious mental conflict to solve, which stimulates access to a distinct resolution, releasing a (tiny) aha). So a different way of putting the explanation for the motive to learn the complexities of syntax, is that doing so is an aesthetically satisfying experience.

(Note that this explanation has a bit of the old behaviorist attempts to unify *why* it is learned with *how* it is learned. An important difference, of course, is the availability of the tool kit and "aha" as the internal "reinforcement" – so it does not explain *what* is learned. NonetheIess, I bet a bit of scholarship will show that this idea also lurked in some behaviorists' writings – they weren't stupid, just in the grips of an iron philosophy of science).

f) Putting parameters back into the acquisition process.

I started this essay noting that parameterization of languages within a particular grammatical framework can provide a useful typological matrix, which can aspire to be constraints on acquisition which narrow the linguistic search space for the child. It is roughly 40 years since the notion of Parameters was first introduced, and 40 years of research have not been univocally kind to the concept. The gradual morphing of syntactic theory from rules to principles to GB and now minimalism have changed the formal language in which parameters are expressed, morphing

the corresponding theories of acquisition – creating different predictions in details. More important, even the descriptive effectiveness of parameters has been seriously questioned: in some cases, the number of parameters has grown via a proliferation of "micro" parameters: the very power of a matrix of parameters has also been its own undoing – a system of just 20 binary parameters generates 4million distinct languages – attempts to deal with this via "macro" parameters, that would make some parameters dependent on the value set for others, have also found serious criticisms. (for reviews see….Newmeyer….etc).

But for our purposes, we can stipulate that some version of parameters is descriptively effective and correct for attested languages. How parameter setting works as an acquisition model is still a problem. Attempts have ranged from prewired "triggers" setting parameters like electric switches, to gradual statistically accumulating settings from a number of similar language experiences. The switch setting model has often been reviled as brittle and a solution only to the logical problem of learning language: the statistical modeling attempts appear to be psychologically more plausible because they are statistical, but that does not in itself make them models of the acquisition process: rather they are only a logical solution to acquiring settings without innate prewired sentential triggers for each parameter.

The suggestion here is to explore how parameters might accumulate and guide the construction of derivations without being directly dependent on sentential input. Rather if a given setting on a parameter leads to an effective derivational solution of an input problem, it can become a constraint/guide to derivational solutions for other superficial conflicts.

At the moment the presentation of the hypothesis about parameters outlined here is free of empirical data. Whether it is free of empirical support awaits investigation.