Four ways from universal to particular: how Chomsky's principles-and-parameters model is not selectionist

David P. Ellerman

Published online: 16 Aug 2016.

Submit your article to this journal ⬀

View related articles ⬀

View Crossmark data ⬀

RESEARCH ARTICLE

Four ways from universal to particular: how Chomsky's principles-and-parameters model is not selectionist

David P. Ellerman

Philosophy Department, University of California, Riverside, USA

ABSTRACT
Following the development of the selectionist theory of the immune system, there was an attempt to characterise many biological mechanisms as being 'selectionist' as juxtaposed with 'instructionist'. However, this broad definition would group Darwinian evolution, the immune system, embryonic development, and Chomsky's principles-and-parameters (P&P) language-acquisition mechanism together under the 'selectionist' umbrella, even though Chomsky's mechanism and embryonic development are significantly different from the selectionist mechanisms of biological evolution and the immune system. Surprisingly, there is an abstract way using two dual mathematical logics to make the distinction between genuinely selectionist mechanisms and what are better called 'generative' or symmetry-breaking mechanisms. This distinction is outlined in this note.

1. Introduction

There has been an enthusiasm for what are generally called 'selectionist' mechanisms as opposed to 'instructionist' mechanisms. This paper uses recent developments in non-classical logic to give a fourfold scheme that provides an abstract but sharp distinction between what might properly be called 'selectionist' mechanisms and 'generative' mechanisms, neither of which is instructionist. The recent development is the logic of partitions, dual (in a category-theoretic sense) to the usual Boolean logic of subsets. Together these two logics provide a simple and abstract way to classify four schemes for 'universals' to give rise to 'particulars'. The principal case in point of a generative mechanism being misclassified by some as selectionist is Noam Chomsky's principles-and-parameters (P&P) model of the language-acquisition faculty or universal grammar (UG).[1]

2. The enthusiasm for selectionist mechanisms

There is a long tradition, growing out of biological thought, to juxtapose 'selectionist' mechanisms with 'instructionist' (or Lamarckian) mechanisms (Cziko, 1995; Dennett, 1995; Edelman, 1987; Jerne, 1967; Medawar, 1960). Originally the distinction was drawn in rather general terms. In an instructionist or Lamarckian mechanism, the environment would transmit detailed instructions about a certain adaptation to an organism, while in a selectionist

---

mechanism a diverse variety of (perhaps random) variations would be generated, after which some adaptations would be selected by the environment but without detailed instructions from the environment.

The selectionist–instructionist juxtaposition gained importance with the development of the selectionist theory of the immune system (Jerne, 1955, 1985). There is some 'generator of diversity' that generates a wide variety of possible adaptations, and then interaction with the environment differentially amplifies some possibilities while the others languish, atrophy, or die off. For instance, in the case of the human immune system, '[i]t is estimated that even in the absence of antigen stimulation a human makes at least 10^{15} different antibody molecules – its *preimmune antibody repertoire*' (Alberts et al., 1994, p. 1221).

This victory for the idea of a selectionist mechanism, not only in biological evolution but also in the immune system, gave rise to an enthusiasm to find many other applications (Cziko, 1995; Dennett, 1995). Gerald Edelman has sharpened the selectionist definition and generalised its application:

> The long trail from antibodies to conscious brain events has reinforced my conviction that evolution, immunology, embryology, and neurobiology are all sciences of recognition whose mechanics follow selectional principles. … All selectional systems follow three principles. There must be a generator of diversity, a polling process across the diverse repertoires that ensue, and a means of differential amplification of the selected variants. (Edelman, 2004, p. 7367)

In particular, Edelman develops a selectionist theory of brain development:

> the theoretical principle I shall elaborate here is that the origin of categories in higher brain function is somatic selection among huge numbers of variants of neural circuits contained in networks created epigenetically in each individual during its development; this selection results in differential amplification of populations of synapses in the selected variants. In other words, I shall take the view that the brain is a selective system more akin in its workings to evolution than to computation or information processing. (Edelman, 1987, p. 25)

The key point is that the possibilities must be in some sense *actualised* or *realised* (e.g. as antibodies in low concentration in the immune system) in order for selection to operate on and differentially amplify or select some of the actual variants while the others languish, atrophy, or die off.

3. Selectionist versus generative approaches to universal grammar

What would the child's language-learning faculty look like if it was selectionist in this sense? There is both a naíve and a sophisticated selectionist account of language acquisition. In the naíve account, the child would (perhaps randomly) generate a diverse range of babblings, some of which would be differentially reinforced by the linguistic environment (Skinner, 1976):

> Skinner, for example, was very explicit about it. He pointed out, and he was right, that the logic of radical behaviorism was about the same as the logic of a pure form of selectionism that no serious biologist could pay attention to, but which is [a form of] popular biology – selection takes any path. And parts of it get put in behaviorist terms: the right paths get reinforced and extended, and so on. It's like a sixth grade version of the theory of evolution. It can't possibly be right. But he was correct in pointing out that the logic of behaviorism is like that [of naíve adaptationism], as did Quine. (Chomsky & McGilvray, 2012, p. 53)

As noted in the quote above, Willard Van Orman Quine adopted essentially this approach to language learning:

> An oddity of our garrulous species is the babbling period of late infancy. This random vocal behavior affords parents continual opportunities for reinforcing such chance utterances as they see fit; and so the rudiments of speech are handed down. … It remains clear in any event that the child's early learning of a verbal response depends on society's reinforcement of the response in association with the stimulations that merit the response, from society's point of view, and society's discouragement of it otherwise. (Quine, 2013, pp. 73–75)

Since language users can generate a variety of rule-based grammatical sentences never before spoken, the child would have to rather miraculously generalise the grammatical rules from the reinforced variants. In order to be adequate as an explanation, a model needs to take seriously the speaker's rule-based competency.

There is a sophisticated version of a selectionist model for the language-acquisition faculty or UG which could be called the *format-selection* (FS) approach (Chomsky, private communication). The diverse variants that are actualised in the mental mechanism are different sets of rules or grammars. Then, given some linguistic input from the linguistic environment, the grammars are evaluated according to some evaluation metric, and the best rules are selected.

> Universal grammar, in turn, contains a rule system that generates a set (or a search space) of grammars, $\{G_1, G_2, \ldots, G_n\}$. These grammars can be constructed by the language learner as potential candidates for the grammar that needs to be learned. The learner cannot end up with a grammar that is not part of this search space. In this sense, UG contains the possibility to learn all human languages (and many more). … The learner has a mechanism to evaluate input sentences and to choose one of the candidate grammars that are contained in his search space. (Nowak & Komarova, 2001, p. 292)

The idea is that after a sufficient stream of linguistic inputs, the mechanism would converge to the best grammar that matches the linguistic environment. Since it is optimising over *sets of rules*, this model at least takes seriously the need to account for rule-based competency. Early work (through the 1970s) on accounting for the language-acquisition faculty or UG seems to have assumed such an approach.

> The earliest ideas were roughly as follows. Suppose that UG provides a certain format for languages, that is, a specification of permitted types of rules and permissible interactions among them. Any rule system satisfying the proposed format qualifies as a possible human language. … The mind employs certain primitive operations to interpret some of the data presented to it as linguistic experience, then selects among the languages consistent with this experience in accordance with an evaluation metric that assigns an abstract value to each language. (Chomsky, 1986, p. 52)

The problems that eventually arose with the FS approach can be seen as the conflict between descriptive and explanatory adequacy. In order to describe the enormous range of human language grammars, the range of grammars considered would make for an unfeasible computational load of evaluating the linguistic experience. If the range was restricted to make computation more feasible, then this would not account for the variety of human languages.

> It was an intuitively obvious way to conceive of acquisition at the time for – among other things – it did appear to yield answers and was at least more computationally tractable than what was offered in structural linguistics, where the alternatives found in structural linguistics could not even explain how that child managed to get anything like a morpheme out of data. But the space of choices remained far too large; the approach was theoretically implementable, but completely unfeasible. (Chomsky & McGilvray, 2012, p.173)

Instead of the FS approach, the alternative *principles and parameters* (P&P) approach (Chomsky, 1981, 1995; Chomsky & Lasnik, 1993; Crain & Lillo-Martin, 1999) to UG was then developed:

> we no longer consider UG as providing a format for rule systems and an evaluation metric. Rather, UG consists of various subsystems of principles; it has the modular structure that we regularly discover in investigation of cognitive systems. Many of these principles are associated with parameters that must be fixed by experience. The parameters must have the property that they can be fixed by quite simple evidence, because this is what is available to the child; the value of the head parameter, for example, can be determined from such sentences as *John saw Bill* (versus *John Bill saw*). Once the values of the parameters are set, the whole system is operative. (Chomsky, 1986, p. 146)

The present purpose is to give an abstract conceptual differentiation of the generative approach (e.g. P&P) from the sophisticated selectionist approach of the FS system (not to mention from the crude selectionism in behaviourism or naïve Darwinism). Some cognitive scientists, such as Piattelli-Palmarini (1989), who agree that UG cannot be accounted for by the selectionism of the FS approach or cruder behaviourist versions, have nevertheless characterised the P&P approach to UG as being 'selectionist' in a much broader sense of being an internal non-instructionist mechanism, a sense in which is better characterised as a generative mechanism (see below).

Jerry Fodor and Massimo Piattelli-Palmarini took a different approach to that differentiation between the P&P treatment of UG and selectionist mechanisms. In the process, they generated some controversy with evolutionary biologists by claiming that 'Skinner's account of learning and Darwin's account of evolution are identical in all but name' – or, to be more precise, 'what is wrong with Darwin's account of the evolution of phenotypes is very closely analogous to what is wrong with Skinner's account of the acquisition of learned behavior' (Fodor & Piattelli-Palmarini, 2010, p. xvi). They emphasise aspects of what is broadly called the 'neo-neo-Darwinism', Evo Devo, or the 'extended evolutionary synthesis' (Pigliucci & Müller, 2010). Instead of wading into that controversy, the approach taken here is to show how selectionist mechanisms (e.g. FS) and generative P&P-type mechanisms can be differentiated at an abstract logico-mathematical level. Hence, it is necessary to turn to a recent development in non-classical mathematical logic.

4. The two dual forms of mathematical logic

George Boole (1854) originally developed what might be called *Boolean logic* as the logic of subsets, not the logic of propositions. The interpretation solely in terms of propositions and the name *propositional logic* came later:

> The algebra of logic has its beginning in 1847, in the publications of Boole and De Morgan. This concerned itself at first with an algebra or calculus of classes, to which a similar algebra of relations was later added. Though it was foreshadowed in Boole's treatment of 'Secondary Propositions', a true propositional calculus perhaps first appeared from this point of view in the work of Hugh MacColl, beginning in 1877. (Church, 1956, pp. 155–156)

When Boolean logic is interpreted as the logic of subsets, variables stand for subsets of some given universe set U, the operations are subset operations, and a *subset-valid* formula or *subset-tautology* is a formula so that no matter what subsets of U are substituted for the variables, the whole formula will evaluate to U for any nonempty U. It is then a theorem (known to Boole), not a definition, that it suffices to consider only the case where $U = 1$ is a
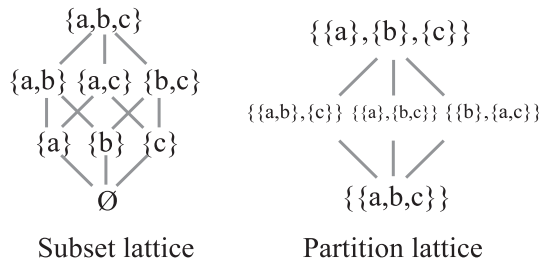
Figure 1. The subset and partition lattices.

singleton which has only two subsets 1 and Ø (the empty set). Hence validity in the special case of propositions with the two truth values 1 and 0, i.e. *truth-table validity*, is equivalent to general subset validity. Eventually, the special case of propositional variables along with the name 'propositional logic' came to dominate, so truth-table validity became the *definition* of a 'tautology' rather than a theorem about subset validity (see *any* contemporary logic textbook).

What is lost by this focus on the special case of propositional logic rather than the general case of subset logic? Around the middle of the twentieth century, the theory of categories was formalised (Eilenberg & Mac Lane, 1945) and an older informal notion of duality in algebra was formalised as the reverse-the-arrows duality of category theory (Mac Lane, 1971). The older informal duality in algebra juxtaposed subgroups with quotient groups, subrings with quotient rings, and, in general, subobjects with quotient objects – which in the basic case of sets was the juxtaposition of subsets with quotient sets (the latter being equivalent to equivalence relations or partitions on a set). For instance, Lawvere and Rosebrugh (2003) call the general notion of a subobject a 'part' and further state that the 'dual notion (obtained by reversing the arrows) of "part" is the notion of *partition*' (p. 85). Hence when the special case of 'propositional' logic is seen as the general logic of subsets, the idea arises of there being a *dual logic of quotient sets or partitions* (Ellerman, 2010, 2014). This idea of a dual logic does not arise when subset logic is seen only as propositional logic, since 'propositions' do not have a category-theoretic dual.

5. The two lattices of subsets and partitions

The two logics of subsets and partitions are represented algebraically by the Boolean algebra of subsets of a universe U and the algebra of partitions on a universe set U ($|U| \geq 2$). For the present purposes, it suffices to consider the two lattices – the familiar Boolean lattice of subsets of U where the partial order is inclusion and the lattice of partitions on U where the partial order is the 'refinement' relation between partitions.

A *partition* $\pi = \{B, B', \dots\}$ on a universe U is a set of nonempty blocks B that are disjoint and whose union is U. Given two partitions $\pi = \{B, \dots\}$ and $\sigma = \{C, \dots\}$ on U, the partition π *refines* the partition σ, written $\sigma \preceq \pi$, if for every block $B \in \pi$ there is a block $C \in \sigma$ such that $B \subseteq C$. Figure 1 illustrates the two lattices for the universe $U = \{a, b, c\}$ (where the partial order is indicated by the lines).

Each lattice has a top and a bottom, and this yields the four universals in the title of this note. In the Boolean lattice of subsets, the top is the universe set U and the bottom is the null set Ø. In the lattice of partitions, the top is the *discrete partition* $\mathbf{1} = \{\{u\} : u \in U\}$ where

all blocks are singletons, and the bottom is the *indiscrete partition* $\mathbf{0} = \{U\}$ with only one block consisting of the universe U.

The duality between the subsets of a set and the partitions on a set extends to the extensive analogies between the *elements of a subset* and the *distinctions of a partition*, where a *distinction* or *dit* of a partition $\pi = \{B, \dots\}$ on U is an ordered pair $(u, u') \in U \times U$ of elements in different blocks of π. There is a set-theoretic representation of the lattice of partitions where each partition is represented by its *set of distinctions* or *ditset*,

$$\operatorname{dit}(\pi) = \{(u, u') : \exists B, B' \in \pi; B \neq B'; u \in B; u' \in B'\},$$

and where the partial order is just inclusion between ditsets since

$$\sigma \preceq \pi \text{ iff } \operatorname{dit}(\sigma) \subseteq \operatorname{dit}(\pi).$$

The complement of a ditset is the set of *indistinctions* of the partition,

$$\operatorname{indit}(\pi) = U \times U - \operatorname{dit}(\pi) = \{(u, u') : \exists B \in \pi; u, u' \in B\},$$

which is simply the equivalence relation associated with the partition. The ditsets of partitions on U are thus the complements of equivalence relations on U and they might be called the *partition relations* on U (also known as 'apartness relations' in computer science).

Given any subset $S \subseteq U \times U$, its *reflexive-symmetric-transitive or rst closure cl* (S) is the smallest equivalence relation containing S (which is well defined since the intersection of two equivalence relations is an equivalence relation). But it might be noted that this closure operation is not a topological closure operation since the union of two rst-closed sets is not necessarily rst-closed. The *interior* int (S) of a subset $S \subseteq U \times U$ is the complement of the rst-closure of the complement, i.e. int $(S) = \left[cl\left(S^c\right)\right]^c$, so it is the ditset of some partition. To define the partition operation corresponding to any logical subset operation (e.g. union, intersection, conditional, etc.), apply the subset operations to the ditsets of the partitions, take the interior of the result, and then take the partition corresponding to that interior. For instance, the meet $\pi \wedge \sigma$ of two partitions π and σ may be defined by the ditset

$$\operatorname{dit}(\pi \wedge \sigma) = \operatorname{int}\left[\operatorname{dit}(\pi) \cap \operatorname{dit}(\sigma)\right].$$

Thus any formula of subset logic can be taken and interpreted as a formula of partition logic. The atomic variables represent partitions on U instead of subsets of U. Given such an interpretation of a formula $\Phi(\pi, \sigma, \dots)$, a member $u \in U$ being an element of the subset represented by $\Phi(\pi, \sigma, \dots)$ is analogous to an ordered pair (u, u') being a distinction of the partition represented by $\Phi(\pi, \sigma, \dots)$. The two definitions of a valid formula are also analogous. A formula $\Phi(\pi, \sigma, \dots)$ is a valid formula of subset logic, i.e. a tautology, if for any subsets of U substituted for the variables, the formula evaluates to the set of all possible elements U (the top of the lattice) for any U ($|U| \geq 1$). Similarly, a formula $\Phi(\pi, \sigma, \dots)$ is a *valid formula of partition logic*, i.e. a *partition tautology*, if for any partitions on U substituted for the variables, the formula evaluates to the partition that makes all possible distinctions, i.e. the top-of-the-lattice discrete partition $\mathbf{1}$ with the ditset $\operatorname{dit}(\mathbf{1}) = U \times U - \Delta$ (where Δ is the diagonal $\{(u, u) : u \in U\}$), for any U ($|U| \geq 2$).

Figure 2 summarises the dual relationships between the two logics (for more details, see Ellerman, 2010, 2014).

	Subset Logic	Partition Logic
'Elements'	Elements u of a subset S	Distinctions (u,u') of a partition π
All 'elements'	Universe set U (all elements)	Discrete partition **1** (all dits)
No 'elements'	Empty set $\varnothing$ (no elements)	Indiscrete partition **0** (no dits)
Duality	Subsets are images $f()$ of injections $f:S{\to}U$	Partitions are inverse-images $f^{-1}()$ of surjections $f:U{\to}T$
Formula variables	Subsets of U	Partitions on U
Logical operations	$\cup, \cap, \Rightarrow,...$	Partition Ops. = Interior of subset ops. Applied to ditsets
Formula $\Phi(\pi,\sigma,...)$ holds at 'element'	Element u is in subset $\Phi(\pi,\sigma,...)$	Pair (u,u') is a distinction of partition $\Phi(\pi,\sigma,...)$
Valid formula $\Phi(\pi,\sigma,...)$	$\Phi(\pi,\sigma,...) = U$ for any subsets $\pi,\sigma,...$ of any U ($|U| \geq 1$)	$\Phi(\pi,\sigma,...) = \mathbf{1}$ for any partitions $\pi,\sigma,...$ on any U ($|U| \geq 2$)

Figure 2. Dual analogies between the subset and partition logics.

6. Four ways to go from universal to particular

In the two lattices of the dual logics, there are four 'universals': the tops and bottoms of the two lattices. The four logico-mathematical ways to abstractly characterise going from universal to particular are the four ways of going from one of the universals to a particular subset or a particular partition in the corresponding lattice. In view of the duality between elements and distinctions, the four ways can be characterised as: (1) killing off elements, (2) creating elements, (3) killing off distinctions, and (4) creating distinctions.

If S represents any particular subset of U and π represents any particular partition on U, then those four ways are:

(1) the *selectionist* mechanism $U \to S$: to go from the universe set U to a particular subset S by 'selecting' the elements of S by eliminating or 'killing off' the elements of the complement S^c;
(2) the *creationist* mechanism $\varnothing \to S$: to go from the empty set $\varnothing$ to a particular subset S by 'creating' the elements of S;
(3) the *classification* or (*symmetry-making*) mechanism $\mathbf{1} \to \pi$: to go from the discrete partition $\mathbf{1}$ to a particular partition π by identifying elements ('killing off' distinctions) of U (in a consistent way); and
(4) the *generative* (or *symmetry-breaking*) mechanism $\mathbf{0} \to \pi$: to go from the indiscrete partition $\mathbf{0}$ to a particular partition π by 'generating' distinctions on U (in a consistent way).

The four schemes can be related in terms of duals (the elements-distinctions duality) and opposites (all versus none), as illustrated in Figure 3.

A selectionist ($U \to S$) mechanism (the top left corner of the square) and a generative ($\mathbf{0} \to \pi$) mechanism (the bottom right corner) are related by taking the dual *and* the opposite (in either order).
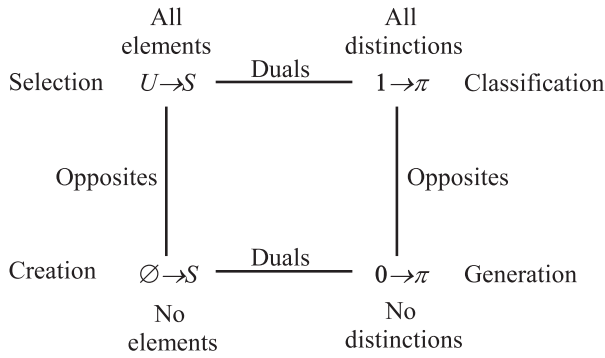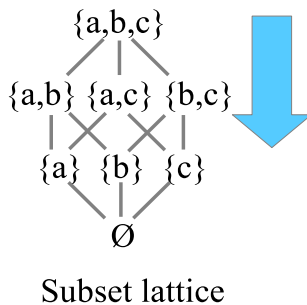
Figure 3. The square of dual and opposite relations between the four schemes.

6.1. The selectionist mechanism: from the universe set to any subset

The first scheme $U \rightarrow S$ (Figure 4) is the abstract logico-mathematical model of a selectionist process, since it starts with an *actualised* set of diverse alternatives U (since selection cannot operate on mere possibilities) and then a number of the alternatives are eliminated by some fitness criterion or evaluation metric while the remaining alternatives are selected (e.g. by differential amplification).

The original example of a selectionist process is Darwinian evolution, wherein the set of diverse alternatives is generated over time by random genetic mutations and then the environment applies a fitness filter (on the general generate-and-filter idea, see Fodor & Piattelli-Palmarini, 2010). From the perspective of the immune system, the actualised set of diverse alternatives is the generated set of antibodies in low concentration and then the selectionist process differentially amplifies those that are effective at eliminating an invading antigen. The various selectionist models of Edelman (1987) also fit this scheme. One of Peter Medawar's (1960) metaphors for a selectionist scheme is a jukebox where all the tunes are already actualised as records in the jukebox and then one is selected. In the FS approach to UG, the mental mechanism must generate some representation of the diverse variety of grammars, and then a chunk of linguistic experience is evaluated according to some evaluation metric to find the best fit among the various systems of rules.

Subset lattice

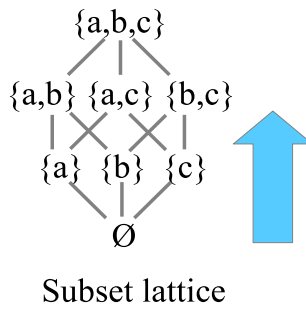**Figure 4.** From the universe set U to a particular subset S.

$$\{a,b,c\}$$

$$\{a,b\} \quad \{a,c\} \quad \{b,c\}$$

$$\{a\} \quad \{b\} \quad \{c\}$$

$$\varnothing$$

Subset lattice

Figure 5. From the empty set $\varnothing$ to a particular subset S.

6.2. The creationist mechanism: from empty set to any subset

The second scheme $\varnothing \rightarrow S$ (Figure 5) is the abstract logico-mathematical model of a creation story, where the elements are, in effect, created out of nothing. While this is one type of 'creation story', the Big Bang creation theory is modelled not by this $\varnothing \rightarrow S$ scheme but by the generative **0** $\rightarrow \pi$ scheme, where the making of distinctions is rendered as symmetry-breaking (Pagels, 1985). The creationist ($\varnothing \rightarrow S$) scheme is perhaps the least interesting to model actual processes, since *ex nihilo, nihil fit*.

6.3. The classification mechanism: from discrete partition to any partition

The third scheme **1** $\rightarrow \pi$ (Figure 6) is the abstract logico-mathematical model of any classification, partitioning (Lawvere & Schanuel, 1997, p. 82), or quotienting process that proceeds by making consistent identifications.[2]

One example is the classification of animals where **1** represents each animal by itself and π might represent the partition of the set of animals as to species. Mathematically, the action of a group on a set is automatically reflexive, symmetric, and transitive so it defines an equivalence relation where the equivalence classes are called 'orbits' (Mac Lane & Birkhoff, 1967, p. 99). This **1** $\rightarrow \pi$ scheme is 'symmetry-making' while the opposite scheme **0** $\rightarrow \pi$ is 'symmetry-breaking'.

$$\{\{a\},\{b\},\{c\}\}$$

$$\{\{a,b\},\{c\}\} \quad \{\{a\},\{b,c\}\} \quad \{\{b\},\{a,c\}\}$$

$$\{\{a,b,c\}\}$$

Partition lattice

Figure 6. From the discrete partition **1** to a particular partition.

$$\{\{a\},\{b\},\{c\}\}$$

$$\{\{a,b\},\{c\}\}\ \{\{a\},\{b,c\}\}\ \{\{b\},\{a,c\}\}$$
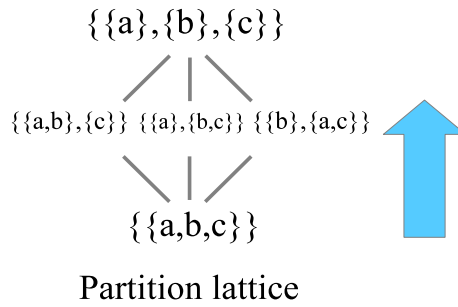
$$\{\{a,b,c\}\}$$

Partition lattice

Figure 7. From the indiscrete partition **0** to a particular partition π.

6.4. *The generative mechanism: from indiscrete partition to any partition*

The fourth scheme $\mathbf{0} \to \pi$ (Figure 7) is the abstract logico-mathematical model of any *generative* or *symmetry-breaking* process where a number of different outcomes (represented by the blocks of π) can be generated by consistently making distinctions.[3]

The most 'atomic' type of distinction is a binary partition of the single block $U \in \mathbf{0}$ into two blocks and then the binary partitions can be combined or joined together. The *join* $\pi \vee \sigma$ of two partitions $\pi = \{B, \dots\}$ and $\sigma = \{C, \dots\}$ is the partition of nonempty intersections $B \cap C$. In terms of ditsets, dit $(\pi \vee \sigma) = $ dit $(\pi) \cup$ dit (σ), where the interior is not needed because a union of partition relations is always a partition relation (since the intersection of equivalence relations is always an equivalence relation). And the most 'efficient' binary partition is one that divides the block $\{U\}$ into two equal parts (assuming an even number of elements). The classic example is where U has 2^n elements which can be enumerated using n-place binary numbers. Then U can be divided into two equal parts by the binary partition according to whether the ith binary digit is 0 or 1. The join of those binary partitions for $i = 1, \dots, n$ would go from the indiscrete partition $\mathbf{0}$ all the way to the discrete partition $\mathbf{1}$, so the n equal-binary partitions are Shannon's n *bits* (Ellerman, 2009).

Often a $\mathbf{0} \to \pi$ generative process proceeds not only by joining binary partitions (with not necessarily equal blocks) but by designating one of the blocks as in a game of 20 questions where the block with the *yes* answer to the *yes* or *no* question is designated. In this case, the $\mathbf{0} \to \pi$ process goes not just from the indiscrete partition $\mathbf{0}$ to a particular partition π but from the single block $U \in \mathbf{0}$ to a specific block $B \in \pi$ (like a correct answer in the game of 20 questions) by following the *yes* branches on the binary tree. This is the case of the generative mechanism that is of most interest for the present purposes.

A binary partition with a designated block is just a choice with two options, and it might be represented by a switch with a neutral setting (representing the state before the choice is made) and then two options such as the Left Option and the Right Option.

For instance, consider the example where U has $8 = 2^3$ elements represented by the three-digit binary digits $b_3 b_2 b_1$. There would be three switches for $i = 1, 2, 3$ where the two options for the ith switch are a 0 or 1 in the ith place b_i of the three-digit binary number. Then the initial state is the indiscrete partition $\mathbf{0} = \{U\}$ where all the switches are in neutral. The first switch (controlling the first digit on the right) determines the binary partition with one block having the four elements $b_3 b_2 0$ (Left Option) and the other block having the four elements $b_3 b_2 1$ (Right Option), and so forth for the other two switches. When all three switches have been set one way or the other, this determines the transition from the single
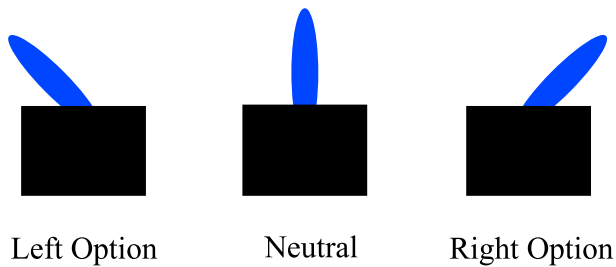
Left Option   Neutral   Right Option

Figure 8. Switch to go from Neutral to either the Left Option or the Right Option.

block $U = \{000, 001, \ldots, 111\} \in \mathbf{0} = \{U\}$ to a specific singleton block represented by a specific three-digit binary number such as $\{010\}$.

The principal application here of the $\mathbf{0} \to \pi$ generative[4] or symmetry-breaking scheme is Chomsky's P&P description of the language-acquisition faculty or UG (Chomsky, 1981, 1995; Chomsky & Lasnik, 1993; Crain & Lillo-Martin, 1999). In this simple model, the parameters are represented by the switches that can be moved to the left or to the right (from the original setting of neutral) by the child's linguistic experience, and the underlying principles are expressed in the whole set-up, defining the grammatical meaning of the left and right settings.

> A simple image may help to convey how such a theory might work. Imagine that a grammar is selected (apart from the meanings of individual words) by setting a small number of switches –20, say – either 'On' or 'Off'. Linguistic information available to the child determines how these switches are to be set. In that case, a huge number of different grammars (here, 2 to the twentieth power) will be prelinguistically available, although a small amount of experience may suffice to fix one. (Higginbotham, 1982, p. 154)

This imagery implicitly allows for a neutral setting on the switches (sometimes called the 'initial state S_0') since otherwise the original or 'factory' setting of the switches would determine a specific grammar independent of experience.[5]

> Borrowing an image suggested by James Higginbotham, we may think of UG as an intricately structured system, but one that is only partially 'wired up'. The system is associated with a finite set of switches, each of which has a finite number of positions (perhaps two). Experience is required to set the switches. When they are set, the system functions. The transition from the initial state S_0 to the steady state S_s is a matter of setting the switches. (Chomsky, 1986, p. 146)

Another implication of this general type of $\mathbf{0} \to \pi$ model for Chomsky's language-acquisition faculty is the interpretation of the adjective 'universal' in the term 'universal grammar'.

> The switch-settings of the metaphor above are in Chomsky's terminology the 'parameters' defined by universal grammar. Notice that this image underscores the sense in which universal grammar, the initial state of the language-learner, need not comprise an account of what languages have in common – to continue the metaphor, different switch-settings could give rise to very different grammatical systems. (Higginbotham, 1982, p. 154)

Thus, 'universal' does *not* mean a specific grammatical rule common to all languages. For instance, three of the left-neutral-right switches set to neutral is 'universal' for determining *any* of the eight possible binary three-digit numbers $b_3 b_2 b_1$ – where the eight numbers have no digit in the same place in common.
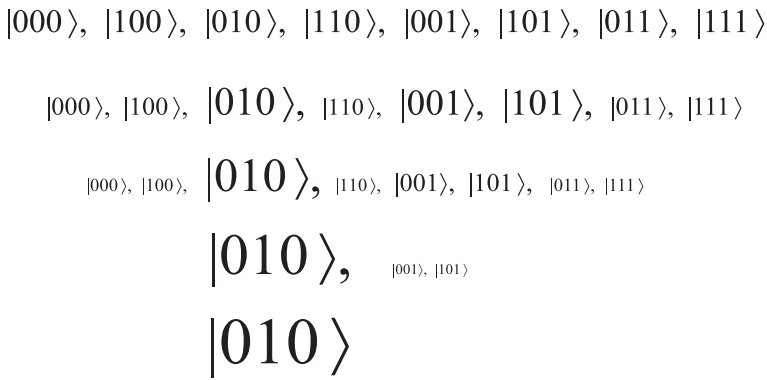
$$|000\rangle, \; |100\rangle, \; |010\rangle, \; |110\rangle, \; |001\rangle, \; |101\rangle, \; |011\rangle, \; |111\rangle$$

$$|000\rangle, \; |100\rangle, \; \Large|010\rangle, \; \small|110\rangle, \; |001\rangle, \; |101\rangle, \; \small|011\rangle, \; |111\rangle$$

$$\small|000\rangle, \; |100\rangle, \; \Large|010\rangle, \; \small|110\rangle, \; |001\rangle, \; |101\rangle, \; |011\rangle, \; |111\rangle$$

$$\Large|010\rangle, \qquad \small|001\rangle, \; |101\rangle$$

$$\Large|010\rangle$$

Figure 9. An abstract model of a selectionist learning mechanism.

Initial State: All switches at Neutral

$$|000\rangle + |100\rangle + |010\rangle + |110\rangle + |001\rangle + |101\rangle + |011\rangle + |111\rangle$$

$$|000\rangle + |100\rangle + |010\rangle + |110\rangle \quad \longleftarrow \quad \begin{array}{l}\text{Left Option at}\\ \text{first position}\end{array}$$

$$\begin{array}{r}\text{Right Option at}\\ \text{second position}\end{array} \longrightarrow \quad |010\rangle + |110\rangle$$

$$|010\rangle \quad \longleftarrow \quad \begin{array}{l}\text{Left Option at}\\ \text{third position}\end{array}$$
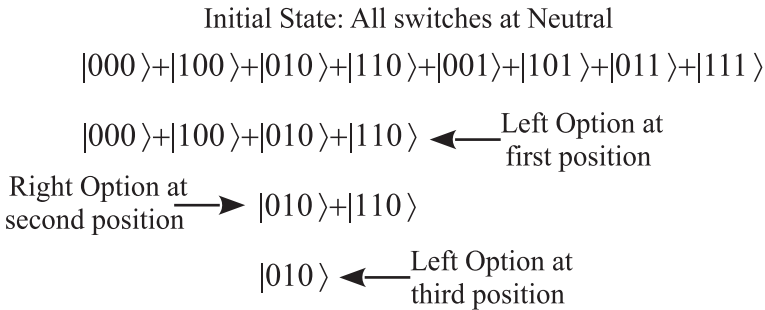
Figure 10. An abstract model of a generative learning mechanism.

7. Generative versus selectionist mechanisms

The two main candidates for biological mechanisms are the selectionist and generative ones, so it might be useful to illustrate a selectionist and a generative mechanism in solving the same problem of determining one among the $8 = 2^3$ options considered in the last section. The eight possible outcomes might be represented as:

$$|000\rangle, |100\rangle, |010\rangle, |110\rangle, |001\rangle, |101\rangle, |011\rangle, |111\rangle.$$

In the selectionist scheme, all eight variants are in some sense actualised or realised in the initial state S_0 so that a fitness criterion or evaluation metric (as in the FS scheme) can operate on them. Some variants do better and some do worse, as indicated by the type size in Figure 9. Eventually the 'unfit' options dwindle, atrophy, or die off, leaving the most fit option $|010\rangle$ as the final steady state S_s.

In the generative learning scheme, the initial state S_0 is where all the switches are in neutral so that all eight potential outcomes are in a 'superposition' (between left and right) state, indicated by the plus signs in Figure 10. It is assumed that the initial experience sets the first switch (or the first parameter in the P&P model) to the left option, which reduces the state to $|000\rangle + |100\rangle + |010\rangle + |110\rangle$ (where the plus signs in the superposition of these options indicate that the second and third switches are still in neutral). Then, subsequent experience sets the second switch to the right option and the third switch to the left option.

Thus, the same outcome of $|010\rangle$ is reached as the final state S_s in the two models, but by quite different mechanisms.

Thus, the above examples differentiate between a selectionist ($U \rightarrow S$) mechanism and a generative ($\mathbf{0} \rightarrow \pi$) one, albeit at a very abstract logico-mathematical level.

There is a whole literature where 'selectionist' is interpreted very broadly as being an internal non-instructionist mechanism so that the generative ($\mathbf{0} \rightarrow \pi$) mechanism is *also* described in those overly broad terms as being 'selectionist' (e.g. by describing the generative mechanism as 'selecting' switch settings):

> I suggest that some important lessons for linguistics and cognitive science can, indeed, be drawn from contemporary biology, but that the new principles and the new assumptions came to bury learning by instruction and to replace it with learning by selection, a radically different process. What now replaces learning everywhere in biology has nothing to do with a transfer of structure and everything to do with mechanisms of internal selection and filtering affecting a pre-programmed chain of multiple internal recombinations and internal 'switches'. (Piatelli-Palmarini, 1989, p. 3)

Now it can be seen that this sort of switch-setting is better described as a generative ($\mathbf{0} \rightarrow \pi$) mechanism, whereas many of the other 'learning' mechanisms in biology (e.g. in the immune system) are correctly described as selectionist ($U \rightarrow S$), since they involve the actualisation of some 'universal' repertoire of possibilities, only some of which are selected. The work of Fodor and Piattelli-Palmarini (2010) might be seen in this light as emphasising the presence of generative 'switch-setting' ($\mathbf{0} \rightarrow \pi$) mechanisms in evolution, as opposed to the accounts that feature only the naïve selectionist ($U \rightarrow S$) mechanisms.

A hierarchy of genetic switches, as in a stem cell or in embryonic development, is a generative mechanism (Jacob, 1973; Monod, 1974). One might imagine a hypothetical selectionist mechanism to replace stem cells that would postulate low concentrations of the different types of cells throughout the body, so that, say, muscle cells would be selected to multiply in a muscle environment while the other types of cells would be inactive there. Yet what is found biologically is *not* that type of selectionist mechanism but the generative mechanism of stem cells (where the muscle environment 'sets the switches' to produce a muscle cell – in addition to reproducing the stem cell).

Medawar (1960) explains the selectionist–instructionist juxtaposition by contrasting a jukebox (with the musical records taken as internal) with a record player (with the records taken as external). The jukebox has a set of pre-existing options, one of which is selected by the simple pushing of a button, whereas when a record player is used to play music, the set of external instructions must be supplied in the form of a record. Thus a jukebox is a genuine selectionist ($U \rightarrow S$) mechanism. Medawar (1982) also describes the development of the embryo as being selectionist: 'Embryonic development … must therefore be an unfolding of pre-existing capabilities, an acting out of genetically encoded instructions; the inductive stimulus is the agent that selects or activates one set of instructions rather than another' (p. 295). But in terms of differentiation for the present purposes, embryonic development is a generative ($\mathbf{0} \rightarrow \pi$), *not* a selectionist ($U \rightarrow S$), mechanism. Medawar (1982) is using 'selectionist' in an overly broad way to describe any internal non-instructionist mechanism. In a similar manner, it is easy to see that Edelman (1987) various models of brain development and learning are all selectionist ($U \rightarrow S$) mechanisms.

In this manner, one could go over all the examples broadly called 'selectionist' and see which are genuinely selectionist ($U \rightarrow S$) mechanisms and which are generative ($\mathbf{0} \rightarrow$

π) mechanisms – which shows the surprising fruitfulness of the quite abstract logico-mathematical differentiation between $U \rightarrow S$ and $\mathbf{0} \rightarrow \pi$ mechanisms.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes

1. As I am a logician and a philosopher of science rather than a linguist, my intent is to set forth the fourfold classification scheme with the possible application to one linguistics model, rather than to take a stand on the many controversies within linguistics.
2. The term 'consistent' means here that the identifications must be reflexive, symmetric, and transitive to form an equivalence relation.
3. In this case, 'consistent' means that nothing can be distinguished from itself, distinguishing must be symmetric, and if u is distinguished from u' and $u = u_1, u_2, \ldots, u_n = u'$, then one of the pairs (u_i, u_{i+1}) must also be distinguished for $i = 1, \ldots, n - 1$, all of which means that the set of distinctions must be anti-reflexive, symmetric, and anti-transitive, i.e. a partition relation (= complement of an equivalence relation).
4. The adjective 'generative' is here used to apply specifically to Chomsky's P&P approach, even though he earlier used it to describe his previous approach.
5. However, it might be the case that *some* 'switches' have only a left or right option and are preset, as it were, to one option. A child would start off with that setting and only switch if a preponderance of evidence indicated the other setting. For instance, in English, 'John didn't eat sushi or pasta' is interpreted to mean only the case 'not-sushi and not-pasta'. But in other languages such as Mandarin Chinese, it is interpreted to mean any of the cases 'not-sushi and not-pasta', 'not-sushi and pasta', and 'sushi and not-pasta', so only the case of John eating both sushi and pasta is ruled out. It seems to be empirically the case that children start off with the parameter setting for the simplest case ('not-sushi and not-pasta'), but when learning Mandarin Chinese (among a number of other languages) they eventually 'flip the switch' to the other setting. (Stephen Crain, personal communication; see also Crain, 2012).

References

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., & Watson, J. D. (1994). *Molecular biology of the cell*. New York, NY: Garland.

Boole, G. (1854). *An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities*. Cambridge: Macmillan.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (1986). *Knowledge of language: Its origin, nature, and use*. New York, NY: Praeger.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, N., & Lasnik, H. (1993). The theory of principles and parameters. In J. Jacobs, A. von Stechow, W. Sternefeld, & T. Vennemann (Eds.), *Syntax: An international handbook of contemporary research* (pp. 506–569). Berlin: de Gruyter.

Chomsky, N., & McGilvray, J. (2012). *The science of language: Interviews with James McGilvray*. Cambridge: Cambridge University Press.

Church, A. (1956). *Introduction to mathematical logic*. Princeton, NJ: Princeton University Press.

Crain, S. (2012). *The emergence of meaning*. Cambridge: Cambridge University Press.

Crain, S., & Lillo-Martin, D. (1999). *An introduction to linguistic theory and language acquisition*. Malden, MA: Blackwell.

Cziko, G. (1995). *Without miracles: Universal selection theory and the second Darwinian revolution*. Cambridge, MA: MIT Press.

Dennett, D. (1995). *Darwin's dangerous idea: Evolution and the meanings of life*. New York, NY: Touchstone.

Edelman, G. (1987). *Neural Darwinism: The theory of neuronal group selection*. New York, NY: Basic Books.

Edelman, G. (2004). Biochemistry and the sciences of recognition. *Journal of Biological Chemistry, 279*, 7361–7369.

Eilenberg, S., & Mac Lane, S. (1945). General theory of natural equivalences. *Transactions of the American Mathematical Society, 58*, 231–294.

Ellerman, D. (2009). Counting distinctions: On the conceptual foundations of Shannon's information theory. *Synthese, 168*, 119–149.

Ellerman, D. (2010). The logic of partitions: Introduction to the dual of the logic of subsets. *The Review of Symbolic Logic, 3*, 287–350.

Ellerman, D. (2014). An introduction to partition logic. *Logic Journal of the IGPL, 22*, 94–125.

Fodor, J., & Piattelli-Palmarini, M. (2010). *What Darwin got wrong*. New York, NY: Farrar, Straus and Giroux.

Higginbotham, J. (1982). Noam Chomsky's linguistic theory. *Social Research, 49*, 143–157.

Jacob, F. (1973). *The logic of life* (B. Stillman, Trans.). New York, NY: Pantheon.

Jerne, N. K. (1955). The natural selection theory of antibody formation. *Proceedings of the National Academy of Sciences of the United States of America, 41*, 849–857.

Jerne, N. K. (1967). Antibodies and learning: Selection versus instruction. In G. C. Quarton, T. Melnechuk, & F. O. Schmitt (Eds.), *The neurosciences: A study program* (pp. 200–205). New York, NY: Rockefeller University Press.

Jerne, N.K. (1985). The generative grammar of the immune system. *Science, 229*, 1057–1059.

Lawvere, F. W., & Rosebrugh, R. (2003). *Sets for mathematics*. Cambridge: Cambridge University Press.

Lawvere, F. W., & Schanuel, S. (1997). *Conceptual mathematics: A first introduction to categories*. New York, NY: Cambridge University Press.

Mac Lane, S. (1971). *Categories for the working mathematician*. New York, NY: Springer.

Mac Lane, S., & Birkhoff, G. (1967). *Algebra*. New York, NY: Macmillan.

Medawar, P. B. (1960). *The future of man: Reith lectures 1959*. London: Methuen.

Medawar, P. B. (1982). *Pluto's republic*. Oxford: Oxford University Press.

Monod, J. (1974). *Chance and necessity: An essay on the natural philosophy of modern biology*. New York, NY: Collins/Fontana.

Nowak, M. A., & Komarova, N. L. (2001). Towards an evolutionary theory of language. *Trends in Cognitive Sciences, 5*, 288–295.

Pagels, H. (1985). *Perfect symmetry: The search for the beginning of time*. New York, NY: Simon and Schuster.

Piattelli-Palmarini, M. (1989). Evolution, selection and cognition: From 'learning' to parameter setting in biology and in the study of language. *Cognition, 31*, 1–44.

Pigliucci, M., & Müller, G. B. (2010). *Evolution: The extended synthesis*. Cambridge, MA: MIT Press.

Quine, W. V. O. (2013). *Word & object (New ed.)*. Cambridge, MA: MIT Press.

Skinner, B. F. (1976). *About behaviorism*. New York, NY: Vintage Books.