

[Special Contribution]**Minimalism:
Where Are We Now, and Where Can We Hope to Go***

NOAM CHOMSKY

*University of Arizona/MIT***Key words:** Enabling Function of Strong Minimalist Thesis, FORMCOPY, FORM-SEQUENCE, M(arkovian)-gap, Minimal Yield

The questions posed have many aspects. I would like to begin with some general remarks on the nature of the inquiry, and then turn to the kinds of problems that arise and what might constitute authentic answers to them. The next step is an effort at principled reconstruction of the basic ideas, followed by extension to new empirical domains, and then a few final reflections.

1. The Nature of the Inquiry

It is useful, I think, to reflect a little on the general background of our inquiry before turning to the specific details of the problems at hand. In particular, to place in a more general context the nature and development of the generative enterprise that began to develop in mid-20th century — guiding concerns that are, not surprisingly, becoming clearer as the enterprise is pursued and surely with surprises to come. I think it may be entering a new phase, one with severe challenges and exciting prospects.

From infancy, children are trying to make sense of the puzzling world around them, and a similar quest is a constant theme of cultural history in all societies we know of. In the modern world, there have been periods where willingness to be puzzled about the world became the spirit of the age, leading to what are sometimes called “scientific revolutions.” One such period was the 17th century. Alfred North Whitehead hardly exaggerated when he wrote that ever since, our “intellectual life ... has been living upon the accumulated capital of ideas provided for them by the genius of the seventeenth century.”

The great breakthrough was that Galileo and his contemporaries allowed themselves to be puzzled by what was just taken for granted in neo-scholastic

* [Editors' Note] This is a revised and extended treatise of the invited lecture, entitled “Minimalism: where we are now, and where we are going,” delivered at the 161st Meeting of the Linguistic Society of Japan, hosted on-line at Tohoku Gakuin University, on November 22, 2020. The video recording of the lecture is available at <https://www.youtube.com/watch?v=X4F9NSVVUw>.

science: rocks fall to the ground and steam rises because they are moving to their natural place; attraction and repulsion are caused by sympathies and antipathies; we perceive some image as a triangle because its form implants itself in our brain; and so on. The new scientists recognized that these were not explanations; rather, descriptive verbiage invoking “occult ideas,” evading the issues. Furthermore, on critical examination, much of what was taken for granted as obvious turned out to be false: for example, that a large ball of lead falls faster than a small one, disproven by Galileo with ingenious thought experiments.

It’s a useful attitude to cultivate, along with other lessons from the genius of this remarkable century.

One such lesson is that thinking hard about very simple properties of nature can be a highly productive feature of inquiry. Why does a ball dropped from the top of a mast of a sailing ship fall to its base rather than behind it as the ship moves forward? — a conclusion that would have quickly been refuted by real-world experiment. It came to be well understood that experiments are radical abstractions from phenomena, theory-driven, highly idealized. Science came to adopt what Stephen Weinberg called “the Galilean style,” giving “a higher degree of reality” to abstract models of the universe than to “the ordinary world of sensations,” a product of too many variables to be informative.

Science also came to adopt the Galilean principle that nature is simple and it is the task of the scientist to show it — a view expressed by many distinguished figures since, in various ways. To mention just one, Herman Weyl observed that “The assertion that nature is governed by strict laws is devoid of all content if we do not add the statement that it is governed by mathematically simple laws ... That the notion of law becomes empty when an arbitrary complication is permitted was already pointed out by Leibniz in his *Metaphysical Treatise* ... The astonishing thing is not that there exist natural laws, but that the further the analysis proceeds ..., the finer the elements to which the phenomena are reduced, the simpler — and not the more complicated, as one would originally expect — the fundamental relations become and the more exactly do they describe the actual occurrences.”¹ A similar thought was captured in a phrase by Jean-Baptiste Perrin in his Nobel Prize address, when he described the essential art of science as reduction of “complex visibles to simple invisibles.”

Some of the most important philosophical inquiry of mid-twentieth century followed a similar path, notably Nelson Goodman’s constructional work, developing ideas of Rudolf Carnap’s, and exploring in depth the intimate relation between theory, explanation, and simplicity. The earliest work in what became the generative enterprise was, in fact, an effort to pursue Goodman’s ideas in the domain of language, an effort that animates much of the enterprise since, taking new forms in the minimalist research program.²

¹ Weyl (1934), cited by Roberts and Watumull (2015). *Discours de métaphysique* (Leibniz 1686).

² For discussion, see Chomsky (to appear).

We have fairly definite concepts of simplicity, but they are not given with precision *a priori*. The concept can be viewed in various ways, and can be refined in the light of consequences of applying it: a reciprocal and mutually supportive relation, not a lethal circularity. The question gains a more substantive interpretation when investigating an organic object: still adopting the Galilean principle, what concept of simplicity did nature use in constructing the simplest solution to problems posed in the course of evolution? Taking human language to be an organic object, the question has immediate significance for inquiry into its nature, and will be at the heart of our concerns here.

These considerations fall within the more general epistemological framework of Duhem–Quine holism, and hence apply as well to choice of Language-Specific Conditions (LSCs), the domain of UG (Universal Grammar). Like all other assumptions, choice of LSCs must be evaluated in terms of empirical consequences, matters that arise below.

Language did not escape the attention of 17th century thinkers. Galileo and the philosopher-logician-linguists of Port Royal expressed their awe and wonder about a truly remarkable fact: with a few symbols, each of us can construct infinitely many thoughts in our minds, and can convey to others with no access to our minds their most innermost workings. For Descartes, the normal creative use of language, employing this capacity, was a foundation for his establishing a second substance, *res cogitans*, relating language and thought: species properties that are at the core of our nature. Galileo himself regarded the alphabet as the most spectacular of human inventions because it captured this marvel. His amazement was echoed in later years in almost the same words by Gottlob Frege, who found “astonishing what language accomplishes. With a few syllables it expresses a countless number of thoughts, and even for a thought grasped for the first time by a human it provides a clothing in which it can be recognized by another to whom it is entirely new.”³

The insight poses what we may call “the Galilean challenge”: How is this achievement possible? The challenge seems to me to capture lucidly the major task faced by the inquiry into the nature of language and mind, into the unique nature of the human species.

Language and thought do seem to be authentic species properties, common to humans in essentials apart from severe pathology and without significant analogue in the non-human world. Our closest relatives, otherwise intelligent apes, cannot begin to grasp the most elementary rudiments of language even with intensive training. They have about the same auditory system as humans, but acquire nothing from the sounds that lead a human infant, almost reflexively, to develop complex systems for constructing and expressing thought. And while other animals are capable of remarkable cognitive achievements, often far surpassing human capacity, none seem to have any remote analogue to human thought, at

³ For discussion, see Chomsky (1966, 2017). Frege cited in Heim and Kratzer (1998) as “the beginning of both symbolic logic and the formal semantics of natural language.”

least in terms of a concept of thought that we grasp with any clarity.

The uniqueness of these properties of the human mind is another strange and puzzling fact. The eminent biologist Ernst Mayr reminded us that there have been some 50 billion species since life emerged on earth almost 4 billion years ago, but “Only one of these achieved the kind of intelligence needed to establish a civilization.” Or in fact even much less: the kind of intelligence needed to freely construct thoughts and to use them to reflect, imagine, conjecture, consider responsibility, and much else common to humans (Mayr 1995).

Mayr’s observation is part of his argument that we may really be alone in the universe: extraterrestrial intelligence may not exist. That might be the solution to Fermi’s paradox: with so many planets like ours in the universe, how can it be that with the most assiduous search, we never find any signs of higher intelligence? Recent studies suggest that within the part of the universe accessible to us under physical law, the probability of life with known mechanisms is “negligibly small,” let alone “intelligent” life (Totani 2020).

Furthermore, this one exceptional creature, *Homo Sapiens*, is a recent arrival on earth, about 2–300,000 years ago. For the previous almost 4 billion years there were, it seems, no thoughts on earth, maybe in the universe. Mayr also reminds us that the appearance on earth of *Homo Sapiens* depended on a series of accidents: a bacterium swallowing a microorganism, leading to the development of complex cells; an asteroid hitting the earth and ending the age of the dinosaurs, enabling small mammals to evolve; and many others. Only at the latest instant of evolution did there emerge creatures with the means to construct and make use of thoughts over an infinite range.

Language and thought, in anything remotely like the human sense, might indeed turn out to be a brief and rare spark in the universe, one that we may soon extinguish. We are seeking to understand what may be a true marvel of the cosmos.

The Port Royal elaboration of the Galilean challenge initiated the rich tradition of General and Rational Grammar, “general” because it sought universal principles underlying human languages, “rational” because it sought to go beyond description to explanation. Throughout, language was generally regarded as closely bound to thought. The common understanding was captured simply in William Dwight Whitney’s phrase that language is audible thought — though we now recognize that sound does not have that unique status. One leading figure, Wilhelm von Humboldt, went still further, identifying language with thought. He also characterized language as “a generative activity” [*eine Erzeugung*] rather than “a lifeless product” [*ein todes Erzeugtes*], *Energeia* rather than *Ergon*, and pondered the fact that somehow this activity “makes infinite use of finite means” — a basic feature of the Galilean challenge.

The intimate association of language and thought also traces back to the origins of the rich tradition of Indian linguistics and philosophy, 2500 years ago. One of its founders held that “language is not the vehicle of meaning or the conveyor belt of thought” but rather its generative principle: “thought anchors language

and language anchors thought ... ‘linguaging’ is thinking and thought ‘vibrates’ through language.”⁴

Humboldt’s formulation of the Galilean challenge brings to light a serious gap in the tradition. It failed to accommodate Aristotle’s crucial distinction in his *De Anima* (Περὶ Ψυχῆς c. 350 BCE) between possession of knowledge and use of knowledge; in contemporary terms, between competence and performance and more specifically, between generation (finitary characterization of the knowledge possessed) and production (a specific action). Humboldt, like those who posed the challenge and the tradition generally, kept to use of knowledge, more specifically, use of knowledge in production. Little was said about perception of language, a prime concern of modern psycholinguistics. Both production and perception access the knowledge that is possessed.

Appropriate means were not available to formulate clearly and investigate the fundamental task: to unearth the system of knowledge possessed, the internal language, I-language in technical usage. It seems that the Aristotelian insight lay dormant until the means were provided in the modern theory of computation created by Turing and other great mathematicians of the early 20th century, and adopted in the generative enterprise.

The study of General and Rational Grammar languished, and was forgotten. In the 20th century it was swept aside by anthropological and structural linguistics. These had major achievements. For the first time, a sharp distinction was made between historical and synchronic linguistics, a major contribution of Ferdinand de Saussure. Careful procedures of analysis were established to determine the elements of language and their arrangement, sharpening many descriptive concepts. And the range and variety of languages investigated was greatly expanded.

By mid-20th century a largely shared consensus had been reached within the small community of professional linguists. As they put it, their new “taxonomic science” rested on two pillars: (i) procedures of analysis and (ii) a conception of the nature and acquisition of language. The procedures yield the elements and their arrangements in each language, these varying with few limits.⁵ The general

⁴ Bhartṛhari. Quoted and paraphrased in Hinzen (2017).

⁵ Claims of virtually limitless variety still persist, based on serious misunderstanding. To cite one typical example, neuroscientist David Kemmerer (2012) writes that “One problem that has not received as much attention as it probably deserves is that, as the language sciences have advanced, it has become harder, rather than easier, to determine which aspects of the uniquely human capacity for language are most likely to have species-typical neurobiological substrates. This uncertainty stems from the fact that extensive typological research on the roughly 7,000 languages in the world has generated increasing evidence that the degree of diversity is far greater than previously assumed (Evans and Levinson 2009). As a consequence, there appear to be few universal properties of language that one could reasonably hope to link with universal properties of the brain.” By “universal properties,” following Evans and Levinson, Kemmerer means shared properties of externalized language, and it is true that on the surface languages appear to vary enormously, just as organisms do. Just as in the latter case, the more we learn about languages the more they seem to share deep proper-

conception adhered largely to Leonard Bloomfield's description of language as "a matter of training and habit." A language is a system of habits, acquired by training and general principles of habit formation, constantly changing with new experience. Anything novel would result from "analogy." Philosophy of language adopted much the same view. The most prominent figure who dealt with these questions, W. V. O. Quine, defined a language as a "complex of dispositions to verbal behavior," established by conditioning. Other conceptions were similar.⁶

The generative enterprise began with skepticism about such assumptions. Reliance on "habit" and "analogy" are wholly at odds with simple observations about language. Furthermore, why should linguistics, alone among the sciences, content itself with taxonomy instead of seeking explanations?

Discontent with the consensus was able to take a concrete form thanks to intellectual tools provided by the new theory of computation. These made it possible to resurrect Aristotle's forgotten distinction between possession and use of knowledge. They also made it clear how a finite mechanism can have infinite output — not resolving Humboldt's puzzle about production, but carving out the domain of feasible inquiry: the study of the knowledge possessed, with finite means and infinite scope. Production, like other creative activity, is beyond reach in any fundamental sense, as is even simple voluntary action.

Most important, these new tools provided the means for developing explanatory theories of a language that (at least) provide "a recursive specification of a denumerable set of sentences," in the words of the first attempt, 70 years ago, to develop what later came to be called generative grammars.

2. Puzzles and Problems

As soon as the generative enterprise was undertaken, problems arose. The taxonomic consensus had no real puzzles. Everything essential was known; the procedures of analysis could be applied to any corpus of materials. But as soon as the search for explanation began, it turned out that almost nothing was known. Virtually every sentence posed new puzzles. Furthermore, the elements that were postulated in the best theories could not possibly be found by a schedule of procedures, just as in every other branch of science. There was an irreconcilable contradiction between the structuralist science and the search for explanation.

Another sharp break was adoption of what was later called the *biolinguistics framework*, adopting Massimo Piattelli-Palmarini's (Morin and Piattelli-Palmarini 1974) proposal: taking language to be a property of the organism, a computational system coded in the brain that for each individual recursively generates an infinite array of hierarchically structured expressions, each formulating a thought, each potentially externalized in some sensory-motor (SM) medium — what we may

ties. For careful analysis of many confusions concerning "universal grammar" and "UG," see Mendiñal-Giró (2018).

⁶ For discussion of this period, see Chomsky (2021a), and from a more personal perspective, Chomsky (2021b).

call the *Basic Property of Language*.

We can think of the analogy of a laptop with a stored program, which can be attached to different printers. The stored program is the analogue of the I-language. It can be externalized by one or another printer without affecting the program. The distinction between I-language and externalization turns out to be more significant than had been realized until recently. We return to that. For now, it's enough to notice that the I-language can be plausibly construed as a system for generation of thought, a modern version of traditional conceptions that have considerable merit, I think.⁷

Adopting this general picture, language is embedded in a network of other biological systems. SM systems access the structures it generates to produce and perceive extra-linguistic events, other cognitive systems access them for reflection, analyzing, and other mental acts. It is conventional to speak of two interfaces, the SM and Conceptual-Intentional (CI) interfaces. While the device is convenient for exposition (and I'll adopt it here), there is no need to postulate the interface levels; access can in principle take place at any stage of the computation.

Explanation has to proceed at two levels. A generative grammar is a theory that seeks to explain the properties of the I-language and the system of externalization possessed by the language user. At a deeper level, the theory of the shared language faculty, UG in modern terms, is concerned with the innate factors that make language acquisition possible — factors that distinguish humans from all other organisms.

UG has goals that appear contradictory. It must meet at least three conditions:

- [A] (i) It must be rich enough to overcome the problem of *poverty of stimulus* (POS), the fact that what is acquired demonstrably lies far beyond the evidence available.
- (ii) It must be simple enough to have evolved under the conditions of human evolution.
- (iii) It must be the same for all possible languages, given commonality of UG.

We achieve a genuine explanation of some linguistic phenomenon only if it keeps

⁷ We might think of the I-language as generating expressions that instruct other cognitive systems, on the analogy of narrow phonetics, which does not in itself determine the sounds of language but rather provides instructions that the articulatory system uses to produce the sounds of speech (and to the perceptual system as to how to interpret linguistic noises, which can be quite different from what is perceived). Analogously, the expressions generated by I-language provide other cognitive systems with the information that they use to construct mental objects that might in some more refined sense be called “thoughts”. See Pietroski (2005), Collins (2019), among others. The concept “thought” fades into obscurity insofar as we depart from linguistically formulated thoughts. That's why Turing, in the famous paper on machine thinking that initiated the field of artificial intelligence, wrote that the question whether machines think is “too meaningless to deserve discussion” (Turing 1950).

to mechanisms that satisfy the joint conditions of learnability, evolvability, and universality, which appear to be at odds.⁸ The course of theoretical inquiry for the past 70 years has been driven by the goal of reconciling these conflicting requirements.

These problems were evident at the outset of the generative enterprise. We now know that they are considerably more severe than what was then envisioned.

With regard to evolvability, genetic studies have shown that humans began to separate not long after their appearance. There are no known differences in Language Faculty, narrowing the window for its emergence. Furthermore, there is no meaningful evidence of symbolic behavior prior to emergence of *Homo Sapiens*. These facts suggest that language emerged pretty much along with modern humans, an instant in evolutionary time. If so, we would expect that the basic structure of language should be quite simple, the result of some relatively small rewiring of the brain that took place once and has not changed in the brief period since.⁹ The apparent contradiction with learnability therefore becomes even sharper.

Research on language acquisition has extended the dilemma further. It has shown that a child of two or three has largely mastered basic properties of its language, including some remarkable ones. Furthermore, statistical study of the data actually available to a child shows that relevant evidence is very sparse (Yang 2013). These discoveries seem to require that UG be very rich in order to overcome the enormous gap between data available and knowledge attained, while the condition of evolvability demands that UG be very restricted. And the problem of apparently endless variation lurks in the background.

3. Addressing the Conundrum: Some Critical Cases

In the past few years, some hope has emerged to resolve this tangle of dilemmas, within the minimalist program. The simplest illustration is a fundamental and quite puzzling universal property of language: structure-dependence.

Consider the properties of (1), mastered by children as young as 30 months of age or less as experiment has shown:¹⁰

- (1) the boy and the girl are/*is in the room

To determine agreement, the child does not use the simplest computational rule, adjacency. Instead, the child reflexively relies on something it never hears: the structure its mind creates. The child then assigns plurality by virtue of the nature of this abstract structure.

⁸ There are further conditions, crucially, neural codability. But too little is known about how the brain handles computation for this to tell us much today. There is even good reason to believe that neuroscience is looking in the wrong place. See particularly the work of C. R. Gallistel, including Gallistel and King (2009).

⁹ For discussion, see Berwick and Chomsky (2016). Dates given there have to be revised in the light of more recent genomic studies about separation of humans, mentioned above.

¹⁰ According to some experimental work, down to 18 months. Shi et al. (2020).

Looking further, consider (2), (3):

- (2) a. the bombing of the cities was criminal
b. the bombings of the city were criminal
- (3) a. the mechanic fixed the car carefully
b. the mechanic carefully packed his tools
c. the mechanic who fixed the car carefully packed his tools
d. carefully, the mechanic who fixed the car packed his tools

In (2a, b), to calculate agreement, the hearer again ignores the simple rule of adjacency and attends to the object constructed by the mind, but must go on to analyze its internal nature, determining the head that assigns θ -roles and label. That's a far more complex calculation than adjacency and in fact involves operations still not fully understood. But there is no option, given UG.

Sentence (3c) is ambiguous. It can be understood as in (3a) or (3b). Sentence (3d) is unambiguous: it means that the mechanic carefully packed his tools, as in (3b). The adverb *carefully* seeks a verb, but it cannot use the simplest computation: pick the linearly closest verb. Like adjacency, the simplest operation is certainly within the cognitive repertoire. A child has no problem picking the first bead on a string. Nevertheless, UG forces us to ignore it and to select the more remote verb — which happens to be the structurally closest verb, as we see when we assign the structure in what our mind tells us is the correct way. It goes without saying that none of this can be learned.

In passing, I should add that there is a substantial literature seeking to show that elaborate statistical study of vast amounts of data can partially approach what the child knows without any evidence — a rather odd effort to begin with, and restricted to the simplest cases — not cases like (3). Others have argued that the result is not surprising because hierarchy is found commonly in nature and even mental life. These efforts collapse on analysis, but more importantly, they are all irrelevant. They are seeking to answer the wrong question. The right question is why simple computations of linear order and adjacency are *not* used, though certainly available. The puzzle is that from infancy on we ignore 100% of what we hear (linear order) and reflexively use only structures that we never hear but that our mind constructs, with non-trivial computations.

There is supporting neurolinguistic evidence. Experimental work designed by Andrea Moro shows that if subjects are presented with invented languages modelled on actual ones, the language areas of the brain react normally; but if the invented languages use very simple rules involving linear order, there is diffuse activity in the brain, indicating that it is being treated as a puzzle, not activating the language areas.¹¹

The reason must be that linear order is simply not available to the I-language, the system that constructs thoughts. We see that in many other cases. Take

¹¹ For discussion, see Moro (2016).

Japanese and English. They are virtual mirror images in basic order, but there is no difference in the thought expressed, say by an SVO or SOV structure. That seems to be true generally, at least for the basic propositional structure of thought.

Why then does speech require linearization? The reason is obvious. The articulatory system cannot produce structures, so the externalization process must impose linear order on an internal system of generation of thought, which is unordered. Sign language is less strictly linear because of wider options available in visual space. The SM systems used for externalization have nothing to do with language; they were in place long before language emerged, and have not changed since. Pure language is the internal system that generates thought.

These observations reinforce the traditional view that language is essentially a system of thought. The modern doctrine that language may have evolved from animal communication seems quite untenable.

These conclusions receive some further confirmation from the fact that structure-dependence interferes with communication. We can see that even from the very simple examples already mentioned. Consider again (1), now with disjunction:

- (4) a. John or Bill is in the room (one or the other)
 b. John or the men *is/*are in the room (either John is in the room or the men are in the room)

There is no way to express the thought intended in (4b) in the simple form corresponding to (4a). If adjacency or linear order were available, there would be an easy solution. This is one of many examples where design of language causes problems of communication, including far more serious ones than this.¹² Quite generally, when there is a conflict between computational and communicative efficiency, the latter is always sacrificed, matters discussed elsewhere. To put it metaphorically, when Mother Nature was constructing language, she was concerned with optimal design, not how the system might be used.

It's worth noting that that's how evolution works quite generally. We can distinguish three steps in normal evolutionary change. The first is innovative; some disruption (mutation, symbiosis, other processes) introduces something new. Then comes reconstruction: nature re-designs the new system in the simplest way, satisfying natural law, matters discussed in depth in classic work of D'Arcy Thompson on laws of growth and Alan Turing on creation of patterns by physical law, more recently in the evo–devo literature. The final stage is winnowing: natural selection reduces the variety to the better adapted. The simplest solution produced at the second stage might turn out to be dysfunctional, but nature doesn't care. In reconstruction, nature seeks optimal design. It has no way to consider possible functions.

¹² The most serious cases involve deletion of copies in accord with computational efficiency, leading to some of the hardest parsing problems. But there are many others. Note that (4b) would be unproblematic for expression of thought if feature valuation kept to late insertion so that only the bare copula reaches the thought level (as in some spoken dialects).

That's why it makes no sense to say that some system evolved *for* X ("the spine evolved for keeping us upright," or "language evolved for communication"¹³).

Language seems to fit the normal pattern. We can envision an evolutionary scenario that looks something like this. Some small rewiring of the brain yielded the general property of recursive generation: the innovative disruption stage. Nature then took the usual course of seeking the simplest such operation (reconstruction), relying on natural law (in this case, principles of computational efficiency), and on a particular concept of simplicity. The result is a system satisfying the Basic Property of Language. Its deepest and most surprising property, structure-dependence, relies on elementary properties of computational efficiency: adopting the simplest computational operation and the No-Tampering Condition (NTC), which basically says "don't add bells and whistles but keep to pure elegance." That's the simplest imaginable condition of computational efficiency. If that makes expressions harder to process and even makes some thoughts impossible to express without circumlocution, too bad. Nature doesn't care.¹⁴

From the properties of structure-dependence (among many others), the only plausible conclusion is that language has two distinct components: the I-language that generates the linguistic formation of thought, and a system of externalization that maps the generated structures to some SM medium. The core language system excludes linear order and other surface arrangements. The external systems require such arrangements, but these are not strictly speaking properties of language: rather, properties of an amalgam of language and systems entirely unrelated to language in their evolution and nature.

In these terms, we can envision a resolution of the conundrum facing UG: condition [A](iii) perhaps can be put to the side in the study of the core system of generation of thought.¹⁵ The variety of languages might be localized in peripheral

¹³ The meaningless statement that language evolved *for* X (communication, etc.) is quite different from the traditional views, reformulated here, associating language and thought. Communication is an action using language (and much else). Language constitutes thought, according to the view discussed here. As Hinzen (*op. cit.*) puts the matter perspicuously, language does not *mirror* an independent thought system; rather language "*is* the sapiens-specific thought system" (with the qualifications of note 7).

¹⁴ Externalization should not be confused with processing. It is pure syntax, part of Aristotelian possession of knowledge. Note that statistical information is irrelevant to I-language as a matter of principle, though as has always been assumed in the generative enterprise (see Chomsky 1957), it can be highly relevant to processing and acquisition — an enterprise with major achievements (see Yang 2016). It can also be relevant to sharpening the meanings of words (as Yang shows), expected insofar as meaning is in part based on use. It is easy to fall into error on these matters, as I did in *LST* (Chomsky 1955) in seeking to determine word boundaries in I-language in terms of transitional probabilities, part of the failure in those early days to distinguish sharply between explanation and procedural analysis of data.

¹⁵ It remains, of course, a major field of inquiry, raising many questions about the nature and structure of the mapping, the role of third factor and economy considerations, and crucially,

aspects of lexicon and in externalization; perhaps completely, we might someday learn. Research seems to me to be tending in that direction, and it would be a natural outcome, though still distant. The POS conditions on the core system are severe, sometimes clearly insurmountable, as in the case of structure-dependence. Furthermore, the SM/I-language amalgam poses a cognitive problem that can be solved in many ways. It would not be surprising to find that it is the locus of the complexity and mutability of languages along with their variety — superficial if these ideas are on the right track.¹⁶

If so, at least for I-language, the conundrum is largely reduced to satisfying the evolvability condition [A](ii). That problem will be overcome to the extent that the structures of I-language are generated by the simplest operations. The *Strong Minimalist Thesis* (SMT) sets this outcome as a prime goal of the theory of language.

Conformity to SMT — hence genuine explanation — is achieved when an account relies on “third factor” principles such as computational efficiency, understood in this context as natural law. These can be taken off the shelf when needed, with no cost. SMT sets as an ideal that all linguistic phenomena can receive genuine explanations in this broader sense. A critical task for inquiry is to see how closely this ideal can be approached.

Returning to structure-dependence, the problem of learnability is overcome if the property is part of UG, as seems clearly to be the case. There is, then, no learning. The problem of evolvability is overcome if the property follows from the simplest combinatorial operation, one that must exist if language is to exist at

the choice and valuation of parameters, hence the core of language variation, it appears. On the latter issue see Roberts (2019), Lightfoot (2020), among others.

¹⁶ The architectural assumption here is a departure from that of the Extended Standard Theory (the “Y-model”) and its early phase-theoretic successors, which held that pure syntax — the “generative engine” in Alec Marantz’s phrase — is supplemented by two interpretive systems: externalization (including phonology and phonetics) and “formal semantics” (FS). Keeping to traditional terms, both are pure syntax, symbolic manipulation — mind-internal, in the biolinguistics framework. Both rewrite the output of narrow syntax in a form that is more perspicuous for language-external systems. The elements they yield as outputs — phones/features, or intensions/events/individuals in a model — are internal objects. Spelling them out is the domain of articulatory phonetics in the former case, the theory of language use in the latter. That could be based on Austinian speech-act theory, including the act of referring, which is not to be confused with the formal relation of reference, the foundation of semantics in the traditional sense, a relation that appears to be lacking for natural language. The two interpretive systems differ crucially in that externalization has many variants and is probably the prime locus of language variation and complexity, while FS is fixed, uniform for all languages, a pure rewriting system, without parametric variation and unlearnable. The conclusion would not have concerned leading semanticists — Tarski, Carnap, probably Frege — whose prime concern was not human language. These are important topics, but outside our concerns here. See Chomsky (1996, 2000, 2013), Pietroski (2005, 2018), among others.

all.¹⁷ In fact, it does. The simplest combinatorial operation is binary set-formation, Merge in contemporary terminology — though the matter is more intricate as we see directly. Structure-dependence follows at once. Linear order is not an option for the child if computation is based on the simplest operation, the one yielded by evolution in its normal course.

Returning to the hypothetical evolutionary scenario, the third winnowing stage may have never been reached, at least for the computational system, perhaps beyond. The reason may be lack of time and variation, or it might derive from the nature of UG if it turns out to be a simple, integrated, principled system that is either all-or-none, in essentials.

I think it is fair to say that the case of structure-dependence is the first genuine explanation of a significant property of language, a surprising and highly significant one in this case.

In these terms, there are genuine Merge-based explanations for other fundamental properties of language beyond structure-dependence. One is the Basic Property of Language itself, a product of Merge-based computation. Another is the ubiquity of displacement with reconstruction (D–R), which had always been regarded (by me in particular) as a curious imperfection of language, a serious and problematic departure from good design, something that has to be explained in some other terms. In fact, D–R follows from the most economical subcase of Merge, Internal Merge (IM), the case that involves least search by a huge margin.¹⁸ The apparent problem has been misconceived. The real question is why language ever resorts to the far richer subcase External Merge (EM). That's clearly a reflection of the fact that language assigns argument structure to generated expressions, which requires EM-generated structures. More generally, language observes Θ -Theory, an LSC.

Reversing the logic, the fact that structure-dependence and D–R exist provides evidence that at least with regard to its core computational properties, language conforms to SMT — and moving a step further, it provides support for the conclusion that language emerged rapidly in evolutionary time, probably along with modern humans.

These observations bring to the fore some general considerations about SMT.

¹⁷ That is as far as the study of language itself can reach in dealing with the evolvability problem. A biological problem of course remains: how did the simplest operation (Merge) evolve? For some ideas, based on work of neurolinguist Angela Friederici, see Berwick and Chomsky (2016). On why Merge couldn't have evolved in steps, see Berwick and Chomsky (2019). For similar conclusions from a different perspective, see Huybregts (2019).

¹⁸ IM with a single lexical item, call it $\{0\}$, yields the successor function and a basis for defining addition. With some further tweaking, it yields arithmetic (in the biolinguistics framework, knowledge of arithmetic). Minsky (1985) found that the simplest Turing machines either crash or yield the successor function (SF), suggesting that SF would naturally evolve, and if some extraterrestrial organism has language at all, it will have SF, a possible base for communication, his concern in this paper. For our concerns here, these observations provide indirect support for the course being pursued.

SMT serves a disciplinary function: it restricts the mechanisms that are available for description of language, a necessity if we are to approach the goal of genuine explanation. Less recognized is the fact that SMT also serves an enabling function. It provides options and systems for language that would have no reason to exist if language did not abide by SMT. Structure-dependence and D-R can be viewed this way. In this regard SMT is somewhat analogous to the laws of form and structure that determine the space of possible organisms, very narrowly it seems.

We will return to some further consequences involving the computational operations and generation of thought, but note first that the enabling function of SMT might extend to the second component of a computational system, the “atoms” of computation.

I-language is a computational system. Any such system has primitive elements and computational procedures to form new elements. In our case, the primitives constitute the lexicon (LEX), atoms of computation for I-language, though (like atoms) they have internal structure that enters into other processes.

LEX is the set of “formatives” in the traditional sense: minimum “meaning-bearing” and functional elements.¹⁹ Investigation shows that even the simplest lexical items (LIs), those used to refer/denote, are highly complex. That much is intimated in the earliest discussions bearing on the matter in classical Greece, for example in Heraclitus’ question how we can ever step in the same river twice. We can reframe the question as a way of exploring the meaning of the word/concept *river*. That turns out to be non-trivial. Thus it is easy to show that under radical changes, R will be taken to be the same river, while under very slight changes it will not be a river at all. These are aspects of the meaning of the LI *river*. Further light on the matter is provided by Aristotle’s discussion of what is a house, again reformulated in cognitive/linguistic terms. In his framework, a house is constituted by *matter* (bricks, timbers, etc.) and *form* (design, intended use, etc.). A one-room school house consists of the same matter whether we refer to it as a house or a room, regarding it quite differently (even more so as a school). These are aspects of the meanings of the LIs/concepts we are using. Much the same is true generally of the simplest words used to refer (see note 7). This must be known without learning, and correspondingly appears to be shared cross-linguistically in fundamental respects. Studies of language acquisition, notably Lila Gleitman’s, have shown that these items are acquired on few presentations, implying that they are based on rich innate substructure. Learning is apparently a matter of settling relatively superficial properties. There are no significant analogues to such concepts in animal symbolic systems and there is no record of their development.

For these reasons, the origin of human concepts has seemed a hopeless mystery, buried deep in the pre-history of Homo Sapiens. But on reflection, that seems most unlikely. It may be another case of the enabling function of SMT. The rich and intricate concepts that are the atoms of LEX would have no function outside

¹⁹ Informal discussion typically speaks of words and their meanings, but *word* of course is a complex concept that is not appropriate in this context.

of a system of generation of thought that yields complex expressions that can enter into reasoning, reflection, and other mental acts. If they had emerged before language, they would have been useless waste and would have been discarded. It is reasonable to suppose that when the capacity emerged for recursive generation of linguistic expressions, it appropriated a lexicon of elementary items available to proto-humans; and only later, as the combinatorial possibilities of generating language and thought became available, concepts of the distinctive human character appeared — how, we can only guess, but at least it might be possible to explore the question if it falls within the history of humans.

Here again the role of SMT might appear. When some innovation appears, nature seeks the simplest way to accommodate it. The simpler the factors that do so, the more refined will be the system that is produced. With many factors intervening the result is likely to be chaotic. It is not much of a stretch of the imagination to think that the enabling function of SMT had a significant role in the development of the unique conceptual resources of human thought, a matter worth contemplating and pursuing. And as more is being learned about the rich, inventive, and creative lives and complex social orders of the tens of thousands of years of human history now becoming accessible to inquiry, new and helpful insights into these questions might be coming into view.²⁰

For present purposes, we put this challenging area of investigation to the side, keeping just to the condition that each LI contains enough information to determine how it is externalized and how it is interpreted by other cognitive systems (its “meaning”).

4. Foundations

Let’s turn now to a principled basis for genuine explanation, adopting SMT as a guideline and asking how closely we can approach the goal.²¹

Our concern here is the computational structure-building procedures for I-language. Keeping to SMT, we assume there is only one operation that forms the expressions satisfying the Basic Property of Language, call it *Merge*. If Nature is observing Galileo’s precept, it should be that *Merge* is the simplest structure-building operation and can be so defined:

[B] *Merge* is the simplest structure-building operation.

Let’s ask how close we can come to this ideal, and what is entailed by success in approaching it.

There is a standard definition of *Merge* that has been used quite effectively, but as we have often found before with proposed mechanisms of computation (phrase structure rules, transformations), it embodies hidden assumptions and permits unacceptable rules. It must be purged of these errors if we hope to achieve

²⁰ See Graeber and Wengrow (2021). Also Mann (2006) on more recent periods.

²¹ Note that SMT is not precisely defined, and can be expected to be sharpened in the course of inquiry. But a general conception suffices for proceeding.

genuine explanations.²² And the ground must be cleared more carefully if this course is to be pursued, a task reminiscent of foundational explorations in science, philosophy, and mathematics.²³

Let's approach the task of a proper formulation of the computational operation for I-language from the ground up, aiming at [B]. As a benchmark, it is useful to consider a standard example of recursive generation of an infinity of expressions without concern for SMT, rules of formation for a formal language, say propositional calculus. Let's spell out details that are always ignored because they are not interesting in this case, though they become significant when we turn to language and SMT.

Take $LEX = \{p, q, r, \dots, \neg, \vee\}$. The rules allow us to form the elements of (5), step by step:²⁴

$$(5) \quad p, \neg p, ((\neg p) \vee q)$$

At each stage of the derivation, we have a set of already generated items that are available for carrying the derivation forward (along with LEX , which is always available). Call this set the *Workspace* (WS). WS determines the current state of the derivation. Derivations are Markovian. The next step doesn't have access to the history, but that doesn't matter here since WS includes everything previously generated.

Suppose $WS = (5)$. The inscription p occurs three times in (5). The computation is "dumb": it sees three different inscriptions that happen to look alike. There is a convention — call it $STABILITY$ — that the three are all *occurrences* of p . The convention is of course necessary for proper interpretation of the derivation. Similarly in a proof, each occurrence of an expression must have exactly the same

²² For the standard definition, clarified and improved, see Collins and Stabler (2016), which also brings out the main illegitimate assumption: an explicit operation that removes the merged elements from the workspace. One of the merits of careful formulation is to highlight illegitimate assumptions. It was soon discovered that removal is necessary (Chomsky 2019a). Subsequent work has sought to overcome this crucial stipulation and other violations of SMT. See Chomsky (2019b), Seely and Kitahara (2021), and Chomsky (forthcoming), which I generally follow here. Much of this work is based on extensive discussions among a group originally formed by the late Sam Epstein, known in the literature as EKS (Epstein, Hisa Kitahara, and Dan Seely), later joined by Robert Berwick, Sandiway Fong, Riny Huybregts, Andrew McInerney, and Yushi Sugimoto. Interactions have been so intense and productive that it would be hopeless to try to identify individual proposals and ideas.

²³ For such reasons, Whitehead and Russell took several hundred pages in their masterwork *Principia Mathematica* to prove that $1 + 1 = 2$. The same is true of the constructional projects of Rudolf Carnap (1926: *Der logische Aufbau der Welt*) and Nelson Goodman (1951: *The Structure of Appearance*).

²⁴ Ignoring further reduction to Sheffer stroke (\uparrow) and Polish notation, eliminating parentheses.

interpretation.²⁵

For simple formal languages, defining *occurrence* is straightforward, using a device of Quine's. For the more intricate language case, it is quite cumbersome, and better dispensed with. That is straightforward. The concept is needed only when several inscriptions are taken to be occurrences of one another. The operative notion is relational. Therefore the notion *occurrence* can be eliminated in favor of a rule FORMCOPY (FC) assigning the relation *Copy* to certain identical inscriptions. One condition on FC is that it observe STABILITY. All of this is trivial, hence ignored, in the formal language case, but has interesting consequences when embedded within SMT.²⁶

SMT holds for organic systems, which face conditions of resource restriction that are not relevant for normal recursion. One crucial condition is computational efficiency, which has many facets, some obvious, such as "do as little as possible" (e.g., NTC). Another natural condition is limiting search. For an operation *O* to apply to items it must first locate them. It must incorporate an operation Σ that searches LEX and WS and selects items to which *O* will apply. It is fair to take Σ to be a third factor element, on the shelf and available for any operation.²⁷

We assume here that minimization of search is Nature's guide in seeking simple solutions for computational systems, the reconstruction stage of evolution discussed above.

Like other efficient operations, Σ seeks to do as little work as possible. We have already appealed to Σ in selecting IM over EM when there is a choice: to apply Merge in the course of a derivation, when possible Σ selects a pair $\langle P, Q \rangle$ with one a term of the other (where *X* is a *term of Y* if *X* is a member of *Y* or a member of a term of *Y*). That vastly reduces search. It is not always possible, however. Language has to satisfy conditions specific to this organic system, LSCs. We have already mentioned one case, Θ -Theory. Language must provide argument structure at CI. That requires recourse to EM. A sharper version is DUALITY OF SEMANTICS: the principle [C], to be sharpened below:

²⁵ The issue has arisen in history of mathematics in a debate over validity of Newton's proofs, at a time when there was no clearly formulated theory of limits. Was there equivocation in his use of zero and "as small as possible"? See Kitcher (1973).

²⁶ Copies need not be marked as such in the derivation. Like other non-structure building operations, FC applies at the phase level and is interpreted (mapped to CI), not entering into further computation. We extend the relation *Copy* to its transitive closure, yielding a notion of *Chain*.

²⁷ There are standard tacit assumptions that are normally put aside, as I will do here, though they should be made explicit and explained in a fuller account. For example, determiners and prepositions are not subject to IM. Here I'll simply refer to the syntactic objects (SOs) that satisfy conditions under discussion as *relevant SOs* — essentially expressions used to refer/denote.

[C] DUALITY OF SEMANTICS:

EM is associated with θ -roles and IM with discourse/information-related functions.

We take DUALITY OF SEMANTICS to be another LSC, by definition restricted to Merge, returning to consequences of adopting the principle.

The LSCs may turn out to be restricted to the core function of language in generating thought, as these and other examples suggest. Note that if Merge satisfies the ideal [B], and really is the simplest structure-building operation, it will automatically provide both EM and IM. The question is why the much less efficient subcase EM is used for language.

Moving a step beyond, suppose the operation O seeks to relate elements P and Q where neither is a term of the other. O will keep to the least search space, the smallest set containing one of them, say P . With P fixed, search for Q is then limited to R , the sister of P in the syntactic object $\{P, R\}$. In this case, P *c-commands* Q . For I-language, it seems that the only step beyond this is “search everything”: EM in the case of Merge.

Another freely available “least effort” condition is *Minimal Search* (MS): Σ searches as far as the first element it reaches and no further. In searching WS , MS selects a member X of WS , but no term of X . In the optimal case of selection of a single syntactic object Z , Σ selects the first relevant term Y of Z , and stops. If Σ is seeking to establish a relation between a member P of Z and some Q , Σ will keep to the *c-command* domain of P , again stopping the search after reaching the first relevant element.

It has always been tacitly assumed that Merge and other operations observe MS, a hidden assumption that we now make explicit as a corollary of limiting Σ (hence of SMT). There is direct empirical evidence for this assumption. Consider (6):

(6) * who_3 do you wonder if who_2 was appointed who_4

Raising of who_2 yields an ECP violation. If minimality of search is abandoned, nothing bars raising of who_1 , which is otherwise a legitimate operation, yielding (6).

Σ is also constrained by the condition of *strict cyclicity*, imposed by *Phase Theory*. Certain SOs are phases. When a phase is constructed, it is dispatched to interpretation at CI and can no longer be accessed by Σ : the *Phase Impenetrability Condition* (PIC).²⁸ Strict cyclicity thus reduces search.²⁹ Choice of phases is not determined by SMT: rather, by LSCs. I will keep here to the assumption that the

²⁸ An assumption that raises questions. There is evidence that Agree is not strictly bound by PIC. See Chomsky et al. (2019).

²⁹ A phase may be unbounded. Accordingly, to compare systems with or without strict cyclicity we resort to the standard device of comparing them for size n (where size might be number of LIs), then selecting the one that is less for all n (past some fixed number). It follows that strict cyclicity is selected as reducing search.

phases are verbal and clausal (v^*P and CP), possibly nominal (nP), for reasons discussed elsewhere.

Merge, we are assuming, is the sole combinatorial operation forming expressions. As such, it must meet conditions that do not hold for non-combinatorial operations. One important condition is that Merge should construct the fewest possible new items that are accessible to further operations, thereby limiting Σ . Merge(P, Q) necessarily constructs one such SO: {P, Q} itself. It should yield no more than that. Call this condition *Minimal Yield* (MY).

Consider IM(P, Q), where Q is a term of P. IM(P, Q) = {P, Q}, where P contains a copy of Q, call it Q'. The operation therefore creates two new elements: {P, Q} and the raised element Q. But Q' is no longer accessible, thanks to MS. Q' is *protected* from Σ by Q. Hence only one new accessible element is added, satisfying MY.³⁰

As we will see directly, MY has substantial empirical consequences. Importantly, it eliminates recourse to the illegitimate operation REMOVE (see note 22). MY is not only natural for organic systems, but is also a plausible candidate for dependence of basic language design on properties of the brain, the first such case of a phenomenon that is very difficult to establish.³¹

We can take the corollaries of SMT just reviewed to determine *Optimal Computation* (OC) for organic systems. The goal of inquiry is to determine how closely the operations of I-language satisfy OC. Thus Merge should satisfy the conditions of Binarity, MS, PIC, restriction to c-command domains, and MY, all corollaries of limiting Σ , itself a property of SMT if our assumptions about Nature's version of simplicity are correct.

We can then take Merge to be the combinatorial operation that satisfies SMT and LSCs. In the most general formulation:

³⁰ The situation is somewhat different with remnant movement, as in (i), where *John* raises by IM to the non- θ subject position followed by raising of [₁ *how likely t' to win*]:

(i) (I wonder) [[how likely t to win][John is [₁ how likely t' to win]]]

Here [₁ *how likely t' to win*] is protected by MS but t' is not, so it seems that two new items are being created, violating MY. However, at least in cases like this, t' is independently protected by PIC, so the problem does not arise. Another possibly relevant consideration is that the remnant movement cases are terminal, not subject to further operations that affect grammatical status or CI interpretation. If so, it wouldn't matter if they violate MY. Many open questions arise.

³¹ See Chomsky (2019b), reviewing ideas of Sandiway Fong's concerning the fundamental neural processes of discarding vast amounts of information provided by super-sensitive sensory organs. For more extensive discussion, see Fong (2021). Note that discarding information — extracting signal from noise — is a sophisticated matter, relying on existing structures that shape and direct the process: sometimes acquired (a radiologist identifying a tumor in an image), sometimes innate (a human infant acquiring a language from data that remains noise to a chimp), always involving still deeper principles of natural law.

[D] $\text{Merge}(X_1, \dots, X_n, WS) = WS' = \{\{X_1, \dots, X_n\}, W, Y\}$, satisfying SMT and LSCs.

Merge therefore is the most economical structure-building operation. It maximally limits search, conforming to SMT and its corollaries. In particular, $n = 2$ and Y is null (by virtue of MY). W is whatever is unaffected by the operation, hence carried over. [D] effectively spells out the ideal [B].

Condition MY has important empirical consequences. Suppose it is violated. Suppose the only members of WS are P and Q . Ignoring MY, Merge yields [E]:

[E] $\{P', Q', \{P, Q\}\}$

[E] permits indeterminacy. If the next operation merges P with X , there are two possibilities, with outcomes differing at the two interfaces. A structure-building operation should not have such a property. Much more seriously, [E] permits legitimate derivations with ungrammatical outcomes. Thus $\{P, Q\}$ can be expanded to R which has arbitrary complexity. P' can be merged with R . The outcome is indistinguishable from P merged with R , which can be hopelessly ungrammatical, as if the object of an unaccusative could raise over arbitrarily many islands.

Plainly, MY must be a condition on Merge. Fortunately, we can take it to be a corollary of SMT. Here again, as in the case of MS, there is direct empirical evidence supporting the conceptually optimal condition, here with a much broader reach. As noted earlier, MY eliminates the need for stipulating REMOVE in the definition of Merge (see note 22).

These considerations rule out all of the suggested extensions of Merge: parallel, multidimensional, sideways, late Merge (which requires other complications as well). The operations all add more than one new element, and unlike IM, no copies are protected by MS. And correspondingly, all are subject to the problems of [E]. The only cases that survive MY are IM and EM, the former much more economical because of the same condition of restriction on search. Accounts using extensions of Merge may be valuable contributions, bringing order out of chaos. But they remain descriptions, perhaps a step towards genuine explanation.

MY renders derivations strictly Markovian in a strong sense, beyond the normal Markovian property of derivations. The derived Workspace, the current state of the derivation, does not contain items that were generated earlier. That is unlike normal recursion, as in proof theory or rules of formation for formal systems, where the history of derivation is contained in the current state. But for language, the situation is different: the history is not preserved in the current state by virtue of MY.

We assume that FC, like other operations, appropriates Σ from the third factor toolkit and operates at the phase level, keeping to MS to select an element X , then searching for a structurally identical element Y under the conditions on Σ , and assigning the relation *Copy* to $\langle X, Y \rangle$.

FC is not subject to conditions that hold for the structure-building operation Merge. We expect, then, to find configurations subject to FC but not Merge,

though at a particular stage of derivation earlier application of Merge is not detectable because of the Markovian property of derivations, which renders history of derivation inaccessible. Call such configurations “Markovian-gaps,” — *M-gaps*. Their existence is predicted by the Enabling Function of SMT, a prediction that is fulfilled, as we see directly.

5. The Enabling Function of SMT

We have already mentioned some examples of the enabling function of SMT. Let’s move on to investigate modules of language that would have no reason to exist if SMT were relaxed.

Merge must satisfy LSCs, including Θ -Theory. One crucial property of Θ -Theory is that assignment of θ -roles is univocal: a single θ -assigner cannot assign two θ -roles to the same element. If that were possible, then the sentence *John saw John*, with two different inscriptions of *John*, could be pronounced *John saw*, taking the predicate-internal subject to be a copy of the object. The verb *see* would then be assigning two θ -roles to *John* (more precisely, to the two copies of *John*).

Generalizing, the core intuitive idea behind Θ -Theory is the principle of univocality, tentatively formulated as [F], where we take a Copy pair to be a special case of a syntactic object SO:

[F] Θ -Theory:

A θ -assigner assigns no more than a single θ -role to an SO and a θ -position cannot receive more than one θ -role.

Copies can be assigned a θ -role only once, but inscriptions with distinct θ -roles can be copies. That possibility is exemplified. Take the sentence form (7):

(7) John tried [X to win]

Suppose X is *John*. In this case, X can delete, unlike *John saw John*. The difference is that there is no Θ -Theory violation. There are two θ -assigners, *try* and *win*.

Application of IM to form (7) would violate DUALITY OF SEMANTICS, but there is no barrier to FC. Hence the $\langle \textit{John}, \textit{John} \rangle$ pair is an M-gap.

To spell it out more fully, for expository convenience, suppose there is an operation *Interpretation* (INT), which surveys the current stage of derivation — that is, the Workspace — and decides what can be done next. Viewing the Workspace, INT can detect the kind of structure created by IM — call it an “IM-configuration” — but lacking access to history, it doesn’t know how it was constructed. INT does however know that identical inscriptions $\langle P, Q \rangle$ in an IM-configuration can be assigned the Copy relation by FC, which differs from IM only in that it is unaware of LSCs. What INT does not know is whether $\langle P, Q \rangle$ was formed by IM or is an M-gap.

The two occurrences of *John* in (7) are assigned θ -roles independently, so the IM-configuration could not have been formed by IM, which is bound by DUALITY OF SEMANTICS. Therefore the IM-configuration must be an M-gap. In this case

M-gaps yield obligatory control — a module enabled by SMT and LSCs (specifically DUALITY OF SEMANTICS).³²

SMT therefore captures Norbert Hornstein's important insight about similarity of raising and control; similarity, but not identity, because there are two kinds of IM-configurations, one derived from Merge and the other from FC: trace and PRO in traditional terms; I'll keep the terms for expository convenience. INT cannot distinguish them, but other operations can, and do. They differ in many ways. For a sample, consider (8):³³

- (8) a. one interpreter each seems [t to have been assigned to the diplomats]
 b. *one interpreter each tried [PRO to be assigned to the diplomats]
 c. l'auteur de ce livre semble [t être intelligent]
 d. l'auteur semble [*en*-t être intelligent]
 e. l'auteur de ce livre croit [PRO être intelligent]
 f. *l'auteur croit [*en*-PRO être intelligent]
 g. la préface risque de n'*en* être pas publiée (trace)
 h. l'auteur n'*en* a pas été persuadé de finir à l'heure (PRO)

There is also performance evidence supporting the distinction between the two kinds of copies. Trace and PRO function differently in perception of language, experimental work has shown.³⁴

The distinction between the two kinds of copy seems well established from several perspectives. It therefore provides empirical support for the assumption that the DUALITY OF SEMANTICS principle [C], on which the distinction rests, is indeed an LSC.

Derivation of the obligatory control module from SMT and LSCs resembles early EQUI-analyses of PRO, abandoned because of a number of problems such as interpretation of (9), analogous to (7):

- (9) many people tried [many people to win]

Though the two occurrences of *many people* are copies in this configuration, their referential properties differ when lacking the Copy relation. That problem is overcome under Phase Theory, with interpretation only at the phase level, where (9) is interpreted as in (10):

³² For similar ideas, developed within an earlier framework, see Martin and Uriagereka (2014). Note that evolution could have made different choices as to the role of Θ -Theory, relying on other versions of simplicity. Thus it could have chosen a stronger notion of univocality: copies must have the same θ -role. That choice would have eliminated control (at least as a system enabled by SMT). Instead, it chose a θ -theoretic condition on Merge (DUALITY OF SEMANTICS), yielding the control module (and more, as we see directly).

³³ Examples (8a, b) are Luigi Burzio's. The cliticization cases are among those discussed by Nicolas Ruwet among many others that involve cliticization. I am indebted to Ian Roberts and Riny Huybregts for the examples.

³⁴ See, among others, Bever and McElree (1988), McElree and Bever (1989).

(10) for many people x , x tried [x to win]

The interpretation, then, is parallel to (7), thanks to the basic principle of Phase Theory [G], with some ambiguities still to be sharpened:

[G] Interpretation is at the phase level.

Not before, not later. “Interpretation” here spells out as access by external cognitive systems. Access at any other stage of the derivation will yield some form of deviance or incoherence.

Suppose X is not strictly identical to *John*, as in (11):

- (11) a. * $[_1 \text{ John } [_2 \text{ tried } [_3 \text{ Mary to win}]]]$
 b. $[_1 \text{ John } [_2 \text{ expected } [_3 \text{ Mary to win}]]]$

(11a) violates the Case Filter. That has been taken to be what motivates raising. We might complicate INT to impose the Case Filter in some fashion. Apart from being unwanted conceptually, the complication raises a variety of ordering and other problems. Suppose again we adhere to the simplest option, adding nothing new.

Case-assignment here is a reflex of the semantic property *Transitivity* (TR) of a verb. To account for (11a), as distinct from the ECM counterpart (11b), it suffices to appeal to TR directly, ignoring the derivative (structural) Case property. Case doesn’t enter into semantic interpretation. Suppose then that Case is part of externalization. Then no problems arise for our immediate purposes, and it doesn’t matter which of the several variants of Case Theory turns out to be correct.³⁵

We might ask why languages have Case systems. There seems to be no general semantic reason. Perhaps establishing relations among elements facilitates perception/parsing.

Idan Landau has raised problems for analyses of the kind just discussed, based on partial control. Compare (12a, b):

- (12) a. John arranged to meet at noon
 b. ??John managed to meet at noon³⁶

The verbs are somewhat similar in meaning, but only *arrange* easily allows the partial interpretation. Compare also the cases of (13):

³⁵ Questions arise about the structure of externalization, ignored here. One question is posed by such expressions as **who did John try who_i to win*. That suggests that in externalization, the Case Filter applies before “third factor” deletion of who_i . A possible mechanism is for INT to mark DELETE on the lower copy. The device may be similar to the narrow-syntactic process of marking DESTRESS on what optionally elides in externalization. There is of course much more to say about all of these matters.

³⁶ Put aside here the irrelevant interpretation *John worked out a way to join the meeting at noon*.

- (13) a. John arranged/??managed for us to meet at noon
 b. John arranged/??managed for them to meet at noon

Interpretations are similar to (12). Perhaps (12a, b) derive from (13a) by deletion of *for us*, with lexical idiosyncrasies, but leaving the basic picture intact.

Deletion of *for us* is found elsewhere. Consider (14):

- (14) a. For Trump to lose the election is impossible
 b. For Trump, for him to lose the election is impossible

Sentence (14a) is ambiguous (with neutral prosody). It might mean that it's impossible that Trump will lose the election or that it's psychologically impossible for him to contemplate that outcome. The latter interpretation presumably derives from deletion of *for him* in (14b). See Epstein (1984).

So far we have been considering raising to subject. Let's turn to raising to object, which should work much the same way:

- (15) a. John expected Bill to leave
 b. John persuaded Bill to leave

These should be analogous to the subject-raising/control cases already discussed.

In (15a), *Bill* raises by IM to SPEC-*expect*. In (15b), *Bill* is merged by EM with the complex verbal structure $\{persuade, \{Bill\ to\ leave}\}$ as its direct object. Within our highly restricted framework, the structures should therefore be as in [H] skipping details, where V is either *expect* or *persuade*:

[H] John INFL $\{ {}_2\ Bill_1, \{ {}_1\ V, \{ Bill_2\ to\ leave\} \} \}$

In both cases, FC determines that the two occurrences of *Bill* are copies. If $V = expect$, the paired Copies were generated by IM; if $V = persuade$, $Bill_1$ was generated by EM, with its own θ -role; it is an M-gap, a PRO structure. In both cases MS determines deletion of $Bill_2$. Thus object control falls into place along with subject control, crucially relying on the distinction between the reincarnation of trace and PRO, now within SMT.

This is a further illustration of the enabling function of SMT. Absent SMT, there would be no reason for the full module of obligatory control to exist.

The analysis of *persuade* resurrects the approach to double complement constructions in *Logical Structure of Linguistic Theory* (Chomsky 1955), now in a far more restricted framework. In the *LSLT* analysis, a transformation formed the verbal complex $[persuade - to\ leave]$, with the object *Bill*. In the analysis here, the infinitive is the complement of *persuade*, and $Bill_1$ of [H] is the object of the verbal complex $[persuade - Bill_2\ to\ leave]$. It would be worthwhile to look into whether double-complement constructions more generally fall into this framework.³⁷

It seems, then, that the obligatory control module is made available by SMT

³⁷ The syntactic object $\{ {}_2\ Bill_1, \{ {}_1\ V, \{ Bill_2\ to\ leave\} \} \}$ in [H] poses a labelling problem that raises further issues about labelling that go beyond the topics discussed here.

and LSCs. We can capture Hornstein's crucial insight about control while avoiding the violation of θ -theoretic univocality. We maintain the PRO/trace distinction with its strong empirical support, along with the LSCs: Θ -Theory and DUALITY OF SEMANTICS.

Let's return to simple transitive sentences, such as *John saw X*. Suppose $X = \text{John}$. With the subject inserted by EM in the predicate-internal position, they are in an IM-configuration. If FC applies, the expression will crash at CI with a Θ -Theory violation. We conclude, then, that like other operations, FC is optional, not applying in this case so that there is no deletion, just two repetitions of *John*.

Another apparent problem is posed by such sentences as (16):

- (16) a. John regards [Bill/himself/*PRO] as a failure
 b. John believes [Bill/himself/*PRO] to be a failure
 c. John expects [Bill/PRO/himself] to be elected
 d. John wanted [Bill/PRO/?himself] to be elected

Such examples indicate that Principle A of the Binding Theory can be taken to be an option of FC — like FC, bound by Optimal Computation (OC) but ignoring Θ -Theory. Further options are determined by lexical properties of the matrix verb, which derivatively determines Case, perhaps in externalization. That much suffices for (16a) and (16b). (16c) requires the option [\pm Transitive] for the lexical entry of *expect*. (16d) shows that for *want* (*hope*, ...) a further condition prefers the weaker option PRO.

The lexical distinctions are illustrated further in such constructions as (17):

- (17) what John [believed/*tried] about Bill

We therefore have a potential approach to incorporation of the fundamentals of Binding Theory within SMT, with a great deal to be filled in and of course leaving many other properties aside.³⁸

In this framework we might also be able to incorporate ideas of Eric Reuland's on univocality of Θ -Theory. In his study of reflexive predicates (Reuland 2014), he suggests that univocality might be reducible to a third factor principle, hence eliminable from UG, by accounting for their basic properties in terms of a general computational principle of *Inability to Distinguish Indistinguishables* (IDI) that reflects "the impossibility for the computational system of human language to handle identicals unless the environment allows them to be distinguished as different occurrences." As he elaborates, "there is no reason to assume that IDI is a principle specific to language. It seems that it is a general property of computational systems that they cannot distinguish different tokens of an expression as different *occurrences* unless there is a work space with minimally either structure or order, enabling this distinction. Thus reflexivity indeed provides us with a window into the way language-specific principles and more general principles interact," a window we can by now perhaps open to further light.

³⁸ On Binding Theory generally, see Reuland (2011).

So far, we have freely adopted the *repetition/copy* distinction among structurally identical elements. The distinction is a long-standing and problematic issue that has been left in some obscurity. Note that contrary to what has been assumed, what is special about language is not the existence of copies, but rather of non-copies (repetitions). Repetitions never appear in general cases of recursion such as rules of formation for formal languages; copies however are ubiquitous, thanks to the uncontroversial principle of STABILITY, already discussed. Hence copies abound, but there are no repetitions.

Repetitions exist in I-language because derivation is in parallel. Thus in an NP–VP structure, NP and VP are generated in parallel, with no interaction, and they might draw independently from the lexicon yielding structurally identical objects that are not copies, as in *John saw John*, with two independent occurrences of *John*. This is not a logical necessity. Evolution might have taken a different course, taking all identical inscriptions to be copies. That’s as far as explanation can reach for now.

Observing STABILITY, Merge produces copies. Because of MY, they only persist in applications of IM. But copies can appear in other ways: as M-gaps, yielding the obligatory control module. The many complications in formulating the distinction have now been overcome. FC, available freely, establishes a Copy relation among certain structurally identical inscriptions. All other such inscription pairs are non-copies (repetitions).

Let’s turn to extending the reach of genuine explanation — from another perspective, the scope of the Enabling Function of SMT.

One elementary case is a long-standing problem about raising. Consider raising to surface subject, with the basic structure [I]:

$$[I] \quad \{ {}_1 C \{ {}_2 EA_1, \{ {}_3 INFL, \{ {}_4 EA_2, vP \} \} \} \}$$

FC applies at the phase level. How then can it determine that the subject EA_1 was raised at an earlier stage of the derivation, before C was merged? This is the norm for A-movement, for both clausal and verbal phases, and has led to considerable complication in earlier work.

We now have a simple answer. FC assigns the Copy relation to $\langle EA_1, EA_2 \rangle$. EA_2 then automatically deletes. There is no need to determine how EA_1 was formed. In fact, due to the preference of IM over EM, and independently the LSC DUALITY OF SEMANTICS, it would have been formed by IM, establishing a Copy relation validated by FC as required in a well-designed system.

The only problem is that the two EAs are unrelated at the phase level where INT applies, and EA_1 lacks a θ -role. However, independently of how EA_1 was formed, it is a copy of EA_2 , which is in a θ -position; EA_1 is therefore θ -linked; a copy is in a θ -position. Suppose τ assigns a θ -role to structural position $P(\tau)$. Then:

$$[J] \quad X \text{ is } \theta\text{-linked to } P(\tau) \text{ if a copy of } X \text{ occupies } P(\tau).^{39}$$

³⁹ Take *Copy* to be reflexive.

Hence both EA_1 and EA_2 are θ -linked to the verbal θ -assigner in [I].

With these notions in place, we can revise Θ -Theory as [K], replacing [F]:

[K] Θ -Theory:

A θ -assigner τ assigns one and only one θ -role to elements θ -linked to $P(\tau)$.

Independently of θ -linking, EA_1 has a semantic role in [I], the most venerable of them all, the role of argument of predication. The effects are exhibited in normal raising as in (18a) or in the existential presupposition of subject position, lacking in 'existential' constructions, as illustrated in (18b, c):

- (18) a. to seem to be intelligent is hard
 b. there is a fly in the bottle/a flaw in the proof
 c. a fly is in the bottle/*a flaw is in the proof

Existential import holds in the same way for raised object:

- (19) John believed/expected a fly to be in the bottle/*a flaw to be in the proof

More generally, the semantics of raising to object is similar to the clausal analogue. Here too the raised object has a semantic role: the option (in fact strong tendency) of *de re* interpretation. In (20a), unlike (20b), *someone* is interpreted *de re*:

- (20) a. John believes someone in the group to be a police infiltrator
 b. John believes that someone in the group is a police infiltrator
 c. The FBI prefers for someone in the group to be a police infiltrator

(20a) means that there is someone in the group, namely Bill, whom John believes to be a police infiltrator. (20b) and (20c) just mean that John believes/the FBI prefers that someone or other in the group is a police infiltrator.

Raising to object has something of the character of what Quine called "exportation" in his influential *de re/de dicto* inquiries, though he wasn't considering infinitives (Quine 1956). And as in the whole *de re/de dicto* literature, there are many qualifications and areas of unclarity of judgment.

The argument roles of subject of predication and of *de re* interpretation are similar, maybe identical at root.

So far the discussion has kept to A-movement, to positions internal to phases and involved in θ -marking. Matters become problematic when we bring in A'-movement, that is, positions at the phase edge.

Consider a simple example of A'-movement, such as (21a) with the structure (21b). Here $X = \textit{Bill}$ and \textit{John}_2 is in predicate-internal subject position. For clarity, let's use boldface for copies that are linked to object θ -role in the intended interpretation.

- (21) a. **Bill**, {John met yesterday}
 b. **X**, {John₁ INFL {**X**₁, {**John**₂ {meet **X**₂ yesterday}}}}

Suppose however that X is structurally identical to *John*. The derivation proceeds exactly as in (21), but the Markovian operation FC will yield the Copy pairs

of (22):⁴⁰

(22) $\langle X, \text{John}_1 \rangle, \langle \text{John}_1, X_1 \rangle, \langle X_1, \text{John}_2 \rangle, \langle \text{John}_2, X_2 \rangle$

Observing these Copy pairs, deletion yields (23a), which appears as (23b) at the SM interface:

(23) a. **John**, { John_i INFL {**John**₁, { John_2 {meet **John** yesterday}}}}

b. ***John**, met yesterday

At the CI interface, however, the derivation should yield (24), analogous to (21a):

(24) **John**, John met yesterday

The problem can be traced to the improper Copy pair $\langle \text{John}_1, X_1 \rangle$, the Copy analogue to improper movement.

The clean and simple way to avoid such problems as these, which abound, is to segregate A- and A'-movement.

[L] From an A-position, FC searches A-positions.

That leaves only one possibility for Copy to hold between the A- and A'-system, namely, when an A'-position c-commands an A-position, specifically the first step of IM, which establishes θ -linking of A'-positions created by raising.⁴¹

Many consequences follow. Consider "tough movement," with the basic structure (25a) illustrated by (25b):

(25) a. X be easy [₁ Y₁ [for John ... Y₂ ...]]

b. many books are easy for John to read

c. John has read many books/many books have been read by John

As we know, [₁ Y₁ [for John ... Y₂ ...]] in (25a) exhibits all movement properties. There has been some debate about how to derive (25b). One possibility is that Y₂ = *many books* raises to Y₁ and then to X (improper movement). Another is that X = *many books* is introduced by EM and that Y₁ = Y₂ is an empty element, so that

⁴⁰ The Copy pair $\langle \text{John}_2, X_2 \rangle$ is barred by Θ -Theory so it is an M-gap, not the result of IM. But that doesn't affect deletion. Since FC is optional, $\langle \text{John}_2, X_2 \rangle$ might be eliminated from (22). Then X and X₂ would be unrelated: they would be repetitions, not copies, as in left dislocation, not the intended interpretation. Other options also fail.

⁴¹ More has to be said about search from an A'-position. I leave that open here. It is tempting to relate this residual case of A/A' interaction to the observation that in successive cyclic movement, there are often visible effects at the intermediate positions while the first CP position reached does not typically exhibit these effects. See Rizzi (1982). The first move establishing segregation has other properties worth thinking about. Consider the example *who did he tell who₁ about who₂*. Either occurrence of *who* can raise, a case of indeterminacy, tolerable only if it leads to no further computational consequences, as here. We return to such cases and the problems they raise.

$[_1 Y_1 [for John \dots Y_2 \dots]]$ is a free variable structure predicated of X and assigning it the role of argument of predication, θ -linked to object of *read* by a straightforward generalization of the notion of θ -linking.

There is direct empirical evidence in support of the latter analysis, with X introduced by EM. In (25c), both active and passive can mean that John is a voracious reader. This interpretation is blocked for (25b), and as (25c) shows, that cannot be attributed to raising. It follows that improper movement must be blocked; the same is true of copying from A- to A'-position, which would form the Copy pair $\langle X, Y_1 \rangle$, X sharing the interpretation of Y_2 even if it were introduced by EM in (25b) ($X = many\ books = Y_1 = Y_2$).

The problem is overcome by segregation of the A- and A'-systems by [L], forcing the correct analysis within the A-system and barring improper movement/copying.⁴²

Segregation of A- and A'-movement is well-motivated on general grounds. The two systems have very different functions within the overall system of generation of thought. A-positions provide the basis for core semantics: θ - and argument-positions. A'-structures are discourse- and information-related. The basic idea that underlies DUALITY OF SEMANTICS.

A'-movement is always to the edge of a phase, a non-argument and non- θ -position. Keeping to this framework, A-movement must be in "one fell swoop." If it were successive-cyclic, the intermediate positions would be anomalous, non-argument/non- θ A-positions.

There is, however, evidence for successive-cyclic A-movement. The strongest evidence comes from Binding Theory. Consider such structures as (26):

(26) John seems to X [Y to appear to Z [$_1$ P to like Q]]

If $X = her$, then neither Z nor Q can be *Mary*, under Condition C. Hence X c-commands Z and Q . But if $X = Mary$, then neither Z nor Q can be *herself*, though each can be *himself*. It seems to follow, then, that a copy of *John* must appear in position Y , requiring successive-cyclic movement.⁴³

One possible escape from this dilemma is to appeal to a distinction between Condition A, which is local, and Condition C, which is unbounded, and to take the underlying structure to be something like (27), as in *To Mary, John may seem to be intelligent, but he really isn't*.

⁴² FC is unproblematic for raising from A'-positions. This is always to the phase edge, so the Markovian property of derivation causes no problems: information is available at the phase level that IM has applied.

⁴³ Without the intermediate raising verb *appear* (thus *John seems to X [$_1$ P to like Q]*) there would be an alternative to successive-cyclic movement. Infinitives might lack EPP so that there is no raising in [$_1 \dots$] and P is SPEC- vP , with vP the infinitival *to-like*. That suffices to yield the binding paradigm. Independently, the *Mary-herself* relation is barred because the subject of the predicate *appear to herself* is *John*, with or without successive-cyclic movement, and that yields an interpretive clash. The point can be perspicuously interpreted in terms of the Reinhart-Reuland concept of reflexive predication (Reuland 2011, 2014).

(27) To X, John seems to appear to Bill [₁ John to like Mary]

With the structure (27), the binding facts follow. We return to a more principled version of (27).

If successive-cyclic A-movement is eliminated, we can reformulate the principle of DUALITY OF SEMANTICS:

[M] DUALITY OF SEMANTICS:

For A-positions, EM and EM alone fills a θ -position.⁴⁴

So far, we have been able to keep to SMT. Let's try to extend the domain of explanation, asking how closely we can approach SMT. Let's consider two hard problems, which have so far resisted analysis: head-movement and unbounded unstructured sequences (UUS's).

Head-movement is a familiar operation, but it is illegitimate. Head-movement is not formulable within the general framework of the generative enterprise; or, as far as I can see, within any framework addressing the conditions for genuine explanation formulated in [A].⁴⁵ Even allowing some illegitimate formulation, it would give rise to other problems. Unlike A- and A'-movement, head-movement typically lacks semantic consequences: for example, interpretation is the same whether a verb raises to INFL or stays *in-situ*. Such differences from normal IM have led to the suggestion that the operation falls within externalization, but that proposal runs afoul of the fact that it appears to be cyclic, an apparent contradiction.

UUS's are deeply problematic as well. They cannot be generated even by the richest rewriting systems or transformational grammars.⁴⁶

Let's then ask what are the minimal assumptions needed to incorporate both problems within the domain of explanatory theory as construed in the biolinguistics framework. I will keep to a general outline, omitting many details.

Let's begin with sequences. We can assume, fairly generally, that wherever there is an XP, it is the limiting case of a sequence $\langle XP, YP_1, \dots, YP_n \rangle$, as in *John ran* as the limiting case of (28) (which can, of course, be embedded at arbitrary depth):

⁴⁴ This can be extended to A'-positions by generalizing θ -role to include positions of the left periphery in Rizzi's sense.

⁴⁵ See Chomsky (2015). Take say V-to-T raising, as in J.-Y. Pollock's analysis of French-English. The result is V adjoined to T, but that is the wrong result; what is intended is T adjoined to V. Furthermore, the rule is unformulable; it violates the basic extension condition. That can be overcome by first raising V to SPEC-T and then amalgamating it with the head T of TP (Ora Matushansky's proposal), the latter a rule with unique properties. It is, however, sufficient just to keep to amalgamation, as we see below, with no additional complications.

⁴⁶ Since unrestricted rewriting systems are universal, they can code the properties of UUS's, but the properties cannot be read off the rules in the intended manner.

- (28) John, Bill, my friends, the actor who won the Oscar, ... ran, danced, took a vacation ...⁴⁷

The order is crucial as we can see by adding *respectively* or other linguistic devices (“in that order,” etc.). For simplicity, let’s keep to conjunction. We therefore need an operation FORMSEQUENCE (FSQ) that selects m members X_i of WS and yields (29):⁴⁸

- (29) $\langle \&, X_1, \dots, X_n \rangle$

The coordinator distributes through the sequence in one or another way in externalization.

There are matching conditions among the X_i ’s, some involving semantic/pragmatic properties, reminiscent of classical zeugma:

- (30) a. John arrived at the hospital [in an ambulance] and [in a coma]
 b. *John arrived at the hospital in [an ambulance and a coma]
 c. ?John arrived at the hospital in [an ambulance and his street clothes]
 d. No Democrat had won Arizona and Georgia since Clinton in [1992 and 1996] respectively
 e. No Democrat had won Arizona and Georgia since Clinton [in 1992 and in an upset] respectively
 f. *No Democrat had won Arizona and Georgia since Clinton in [1992 and an upset] respectively

Interpretation differs for coordination of PP and NP. PP coordination can be interpreted as two independent events; NP coordination describes a single compound event, hence is subject to more strict matching. The distinction becomes clearer in cases like (31):

- (31) a. John arrived early, met Bill, and got a good seat
 b. to arrive early, meet Bill, and get a good seat seems/*seem to be what John wants
 c. arriving early, meeting Bill, and getting a good seat seems/*seem to be what John wants
 d. to arrive early, to meet Bill, and to get a good seat seems/seem to be what John wants
 e. John read a book by [Tom and Bill]
 f. John read a book [by Tom and by Bill]

Case (31a) describes three independent events. Cases (31b, c) describe a single compound event. (31d) can be interpreted either way. (31e) is joint authorship, unlike (31f).

Generation of these structures first selects X_1, \dots, X_m from WS, forming Y

⁴⁷ Grammaticality doesn’t require equal numbers of coordinated subjects and predicates.

⁴⁸ It may be that $m < n$; X_i might appear several times in the sequence (29), as in *John, Mary, and John saw Tom, Jane, and Jill, respectively* (the same *John*).

= $\{X_1, \dots, X_n\}$, freely using the core operation of set-formation already discussed. Merging of $\&$ and FSQ yields $\langle \&, X_1, \dots, X_n \rangle$, where the X_i 's exhaust the elements of Y . The two operations yield, for example, (32) and then, optionally, (33):

- (32) John lived on a farm with his family
 (33) John lived on a farm and with his family

These differ in that extraction is possible in (32) but not (33):⁴⁹

- (34) which farm did John live on ~~which farm~~ with his family
 (35) *which farm did John live on ~~which farm~~ and with his family

Such constructions as (32), which have no bound, are freely available. They are marginal, except as the stepping stones to sequences.

There are matching conditions for both the set and the sequence, but they are much more stringent for the sequence, and in the former case might fall under non-linguistic incongruity, perhaps alongside of what counts as a plausible metaphor or idiom, highly fluid notions.

We may regard the Coordinate Structure Constraint (CSC) as a violation of the more stringent matching constraint on sequences. If this is the correct interpretation of CSC, then extraction should be possible if it applies to all conjuncts, satisfying the matching condition, as in (36):

- (36) which book did John buy ~~which book~~ and read ~~which book~~

Such cases merit closer attention. Compare (37) and (38):

- (37) which farm did John live on [₁ which farm] near [₂ which farm]
 (38) which farm did John live on [₁ which farm] and near [₂ which farm]

In (37), both [₁ *which farm*] and [₂ *which farm*] can be deleted, yielding a grammatical question that is senseless, since John can't live both on and near the same farm. In (38), deletion of both is less acceptable, perhaps ungrammatical.⁵⁰ Suppose in both cases that only [₁ *which farm*] is deleted. Then (38) is ungrammatical by CSC. The case of (37) with [₁ *which farm*] deleted is more complex. Recall that FC is optional. If FC were to apply, INT would have no information as to which occurrence of *which farm* was deleted and it would again be senseless, as when both are deleted. If FC does not apply, then (37) has a sensible interpretation, with different farms in mind.

The conclusion seems paradoxical. How (37) and [₂ *which farm*] are interpreted depends on whether FC applied. But it is natural to assume that in produc-

⁴⁹ Note that raising as in (34) could lead to ambiguity in such cases as *he lived on which farm near which farm*. Judgments are difficult because examples have other problems, as discussed below.

⁵⁰ Recall that *grammatical* is a notion internal to explanatory theory, while *acceptable* is internal to some experimental framework, theory-laden like all observations but in a totally different sense.

ing the sentences, speakers first decide whether or not distinct farms are in mind, and then apply the rules accordingly; the thought precedes the utterance and determines its form.

The paradox is resolved by distinguishing generation from production. Generation is a matter of possession of knowledge (competence); production is use of knowledge (performance). About the latter, we know very little though it seems reasonable to suppose that the thought precedes the utterance. But these concerns are foreign to generation, which simply lays out the terrain: here is the infinite set of objects — linguistically-formulated thoughts — that are available for use in performance (along with many other factors).

With this much background, consider one of the more complex sequences, pairing unaccusative and transitive, keeping to essentials:

(39) John arrived and met Bill.

The first step is to form the two independent objects $\{ {}_1 v, \{arrive, John_1\} \}$ and $\{ {}_2 John_2, \{v^*, \{meet, Bill\} \} \}$ of (40a) (*John arrived* and *John met Bill*) in the normal way, satisfying Θ -Theory. The next step is to construct the *set* (40b):

- (40) a. $\{ {}_1 v, \{arrive, John_1\} \}, \{ {}_2 John_2, \{v^*, \{meet, Bill\} \} \}$
 b. $\{ \{ {}_1 v, \{arrive, John_1\} \}, \{ {}_2 John_2, \{v^*, \{meet, Bill\} \} \} \}$

I am adopting here the Borer–Marantz root-categorizer analysis, so that *arrive* and *meet* are roots interpreted as verbal with the categorizer *v*. The *v/v** and phasal distinction is determined by the lexical content of the root, with no need to postulate different categorizers. I will distinguish them only for expository convenience.

Application of Merge, introducing $\&$ and INFL, yields (41):

- (41) INFL, $\{ \&, \{ \{ {}_1 v, \{arrive, John_1\} \}, \{ {}_2 John_2, \{v^*, \{meet, Bill\} \} \} \} \}$

Since we so far have only sets, extraction is possible, so either of the occurrences of *John* can be raised to SPEC–INFL, yielding (42), which is converted to (43) by merging C and then applying FSQ:⁵¹

- (42) $John_3, \{ INFL, \{ \&, \{ \{ {}_1 v, \{arrive, John_1\} \}, \{ {}_2 John_2, \{v^*, \{meet, Bill\} \} \} \} \} \}$
 (43) $C, \{ John_3, \{ INFL, \langle \&, \{ {}_1 v, \{arrive, John_1\} \}, \{ {}_2 John_2, \{v^*, \{meet, Bill\} \} \} \rangle \} \}$

Either $John_1$ or $John_2$ can raise by IM to the position of $John_3$. The one that does not raise is determined by FC to be a copy of $John_3$. INT cannot determine which one raised. But it doesn't matter. The choice doesn't affect later operations: there is no indeterminacy. There is also no ambiguity: the interfaces are not affected. Whichever choice is taken, what reaches CI and SM is the same.

As in the case of Control Theory, interpretation of sequences relies crucially on the M-gap. This again illustrates the Enabling Function of SMT. If language

⁵¹ The operation is not cyclic, but that is basically no different from valuing an embedded feature at the phase level. Since this is all phase-internal, PIC does not apply.

were not constrained by SMT, these constructions would have no reason to exist.

In the example (39), the two conjuncts share tense, but that is not necessary. Thus the same analysis applies to (44):

(44) John arrives every day at noon and met Bill yesterday

It also applies with distinct aspectual and modal structures that we're not considering here. It follows that tense is a feature of *v*, not of INFL, contrary to what has long been assumed (the reason why I have been using the older notation *INFL* throughout instead of the conventional *T*). Φ -features however are fixed for the conjuncts, constituting INFL.

Notice that ATB falls out as a special case, with (45c) formed from (45a)–(45b) in a manner analogous to (43):⁵²

- (45) a. (I wonder) [{John bought *what*₁}, {Bill handed *what*₂ to Tom}]
 b. (I wonder) [*what*₃, C, <&, {John bought *what*₁}, {Bill handed *what*₂ to Tom}>]
 c. (I wonder) *what* John bought and Bill handed to Tom

As before, either *what*₁ or *what*₂ can raise to *what*₃ before FSQ applies; whichever one does, the other can be assigned the Copy relation by FC, imposing strict identity.⁵³ INT cannot determine which in fact raised, but again there is no lethal ambiguity.

Suppose we used the same procedure to derive (46a) from (46b):

- (46) a. *(I wonder) *what* John bought and Bill handed a sandwich to Tom
 b. (I wonder) [*what*₁, C, <&, {John bought *what*₂}, {Bill handed a sandwich to Tom}>]

FSQ induces a violation of CSC, a matching violation.

Among other questions to be explored are what appear to be parallelism conditions, illustrated for example in (47):

- (47) *(I wonder) [*who*₃ [[₁ John met *who*₁] and [₂ *who*₂ insulted Bill]]]

The problem is overcome in cases like (48), as observed by Williams (1978):

- (48) (I wonder) [*who*₃ [[₁ John met *who*₁] and [I think [₂ *who*₂ insulted Bill]]]]]

The most straightforward explanation for the difference seems to be the Vacuous Movement Hypothesis (VMH) of Chomsky (1986).⁵⁴ In (48), both *who*₁ and *who*₂ can raise to *who*₃. But in (47), *who*₂ remains in SPEC-INFL, not raising.

⁵² This is a reformulation within copy theory of the analysis of ATB in Huybregts (2018), based on an “equivalence” principle with interesting consequences that he discusses.

⁵³ This raises again the apparent paradox discussed in connection with (37), with the same resolution.

⁵⁴ VMH has to be reformulated within this framework, most simply as a condition on position of Q. I'll put that aside here.

Hence CSC is violated, with *wh*-raising only in conjunct [₁ *John met who*].

Another system that might be enabled by M-gaps is the parasitic gap (PG) construction. Examples (49) illustrate the basic properties of the construction:

- (49) a. what₁ did John₁ file what₂ [without [what₃ John₂ reading what₄]]
 b. John₁ wrote a memoir [without John₂ once referring to himself]
 c. *what₁ was filed what₂ [without [what₃ [John reading what₄]]]

(49a) is a simple PG construction. (49b) illustrates the fact that MS reaches into the adjunct, so there is no barrier to (49a). (49c) follows from VMH, which prevents raising of *what*₁, and segregation of A-/A'-movement, which bars the improper Copy pair <*what*₁, *what*₃>. ⁵⁵

Note that a two-membered sequence is a pair, so it can have an interpretation as asymmetric Pair-Merge, an operation that seems appropriate for many circumstances. Thus *young man* has the properties of *man*, not *young*. The same analysis seems appropriate for (27), repeated here:

- (27) *To X*, John seems to appear to Bill [*John* to like Mary]

The italicized phrase *To X* might be in a separate dimension, pair-merged to *seem*. ⁵⁶

There are other plausible candidates for pair-merge, which also has implications for interpretation of sequences. Interesting matters that would take us too far afield here.

There is a good deal more to say about all of these topics, but this seems to me to capture much of the essence of the matter. If so, the phenomena fall within the domain of explanatory theory, with a limited departure from SMT: FORMSEQUENCE (FSQ). That much seems unavoidable given the empirical facts, and may not be a departure at all, if the operation FSQ can be regarded as part of the “third factor” toolkit.

Many open questions arise, among them the effects of recasting all previous analyses as the limiting case of sequences, exploring the effects of different coordinators and of sequences within sequences, and the very intricate questions involving interpretation — how intricate is illustrated by Barry Schein’s monumental study of conjunction (Schein 2017).

The discussion so far has made no resort to head movement. In fact, it provides a way to eliminate that unformulable operation. For interpretation by INT — mapping to CI — structures such as (50) suffice:

- (50) C, {John₃, {INFL, <&, {₁ *v*, {arrive, John₁}}, {₂ John₂, {*v**, {meet, Bill}}}>}}

⁵⁵ The same holds for * [₂ *what*₁ did you say [₁ *what*₂ was filed [without reading]]] if raising of *what*₂ does not pass through SPEC-C of [₁ *what*₂ was filed [without reading]]. Note that the PG interpretation is assigned at [₁ *what*₂ was filed [without reading]], not [₂ *what*₁ did you say [₁ *what*₂ was filed [without reading]]]; *file without reading*, not *say without reading*. Many questions arise that go beyond what we are considering here.

⁵⁶ Similarly, *to Bill* could be pair-merged to *appear*.

Externalization, however, requires an operation that amalgamates the inflectional elements INFL and *v* along with the roots, presumably cyclically. The operation AMALGAMATE forms a complex [INFL, [*v*, root]]. There are further effects with some variation, as is characteristic of morphology.

The account could be extended to V-to-C raising, but that seems to me a questionable move. This step may fall under the V-second phenomenon, a very different matter. It is restricted to root sentences and embedded structures interpreted as roots. Furthermore, more complex elements can raise, as in *what have been the main reactions to these revelations*. I think these phenomena should be put aside for further inquiry.

The first step in a derivation must select two items from the lexicon, presumably a root R and a categorizer CT, forming {CT, R},⁵⁷ which undergoes amalgamation under externalization, possibly inducing ordering effects that are excluded from I-language in the sense adopted here.

With head-movement eliminated, *v* need no longer be at the edge of the *v*P phase, but can be within the domains of PIC and Transfer, which can be unified. EA is interpreted at the next phase. The conclusion seems more natural than previous versions of Phase Theory.

If this approach is viable, then both of the initial problems dissolve: head-movement is eliminated from I-language; and at most a minimal departure from SMT suffices for the many problems of UUS's.

More generally, it seems that genuine explanations can be provided for an interesting range of core properties of I-language, keeping to search procedures satisfying optimal computation and operations that can plausibly be assumed to be freely available in accord with SMT, along with general principles of economy and computational efficiency that can be regarded as natural law. That carries us forward on the course of determining how closely I-language approaches to SMT, a significant discovery to the extent the conjecture proves valid.

It also seems that for the first time we can see some light in resolving the fundamental dilemma that arose when undertaking the biolinguistics program — that is, when facing the effort to study human language as part of the biological world. The effort faces the conundrum [A], repeated here:

- [A] (i) UG must be rich enough to overcome the problem of POS.
 (ii) UG must be simple enough to have evolved under the conditions of human evolution.
 (iii) UG must be the same for all possible languages.

⁵⁷ The assumption so far is that the categorizers are *n*, *v*, *a*, *p*. Perhaps, as has been suggested, these are analyzed further as [\pm substantive], [\pm predicate]. Perhaps phases are the sets with unmatched values ($+n, -v$), ($-n, +v$), along with CP (with force indicator and left periphery). There is a parallel that might be explored between specific NPs and transitive VPs, both barring extraction. And many other questions that I will put aside.

Merge-based systems relying on third factor properties show considerable promise in approaching the goal of reconciling [A](i) and [A](ii). And there seems to be hope in approaching [A](iii), with UG restricted to language proper (I-language), not its interaction with non-linguistic systems employed for externalization.

It might even be possible to aim higher. As we have seen, we might think of SMT not just as a constraint on what can appear in language (which it is), but as a facilitator of the richness of human language. A different perspective, posing new questions, some explored above.

The projects that lie at the horizon are extremely ambitious, and not long ago would have been regarded as hopelessly premature, if even worth considering. Much less so now. It might be that the ancient study of language is moving to a new era, with new questions and exciting prospects.

6. Some Concluding Thoughts

We might finally ask how the biolinguistics program finds its place in the long and rich history of the study of mind, briefly alluded to in the opening comments. That's plainly too large a topic to try to deal with seriously here, but there are a few guideposts. At the most general level, we can think of the injunction of the Delphic Oracle to "know thyself," a call that we can justly reinterpret in collective rather than individual terms in light of the shared cognitive capacities that so sharply distinguish humans from the rest of the known organic world. A central part of the study of mind should be addressing "the Galilean challenge," at least those aspects of it that fall within our capacities: to account for the human ability to construct from a few symbols an infinite variety of thoughts that constitute our mental lives, and to make use of them to convey to others much of the inner workings of our mind, abilities that rightly evoked awe and amazement in the early days of the modern scientific revolution. Pursuing this path, we can revive the critical Aristotelian distinction between possession and use of knowledge, and with it the distinction between generation and production, long obscured.

A defining feature of the biolinguistics program is taking seriously what is known in the history of philosophy as "Locke's suggestion," the idea that thought is a property of some kind of organized matter, an idea that arose quite naturally after Newton had undermined classical mind-body dualism by showing that there are no bodies in the only sense that was understood. As Darwin captured the common understanding of the time, "Why is thought, being a secretion of the brain, more wonderful than gravity, a property of matter?"

We accept the latter and seek to understand its properties. Why not the former?⁵⁸

Locke's suggestion was developed through the 18th century, most fully in the work of chemist-philosopher Joseph Priestley. It was gradually forgotten. In the 20th century it was largely displaced by behaviorist and structuralist currents.

⁵⁸ For discussion and sources, see Chomsky (2016).

The generative enterprise, within the developing biolinguistics framework, unwittingly returned to Locke's suggestion. The same was happening in other areas. Towards the end of the century, Locke's suggestion was rediscovered (again unwittingly) as the thesis of the new biology that "Things mental, indeed minds, are emergent properties of brains" — neuroscientist Vernon Mountcastle's formulation of the prime insight of the Decade of the Brain that ended the 20th century. Others proclaimed the "astonishing hypothesis" of the new biology, a "radical" new idea in the study of mind, "the bold assertion that mental phenomena are entirely natural and caused by the neurophysiological activities of the brain," and so on. In fact, all reiterate, in virtually the same words, formulations of centuries ago.

My feeling is that it is essentially in those terms that the biolinguistics framework of the generative enterprise should be understood, including the current quest to determine to what extent the human capacity for language and thought actually falls within some plausible version of the Strong Minimalist Thesis.

References

- Berwick, Robert C. and Noam Chomsky (2016) *Why only us: Language and evolution*. Cambridge, MA: MIT Press.
- Berwick, Robert C. and Noam Chomsky (2019) All or nothing: No half-Merge and the evolution of syntax. *PLoS Biology* 17(11): e3000539. <https://doi.org/10.1371/journal.pbio.3000539> [accessed June 2021].
- Bever, Thomas G. and Brian McElree (1988) Empty categories access their antecedents during comprehension. *Linguistic Inquiry* 19: 35–43.
- Chomsky, Noam (1955) *The logical structure of linguistic theory*. Unpublished manuscript, MIT. (1956 revision published in 1975 by Chicago: The University of Chicago Press/ New York: Plenum Press.)
- Chomsky, Noam (1957) *Syntactic structures*. London: Mouton.
- Chomsky, Noam (1966) *Cartesian linguistics: A chapter in the history of rationalist thought*. New York: Harper & Row. [Third edition, with fuller and improved translations, introduced and edited by James McGilvray, 2009. Cambridge.]
- Chomsky, Noam (1986) *Barriers*. Cambridge, MA: MIT Press.
- Chomsky, Noam (1996) *Powers and prospects: Reflections on human nature and the social order*. Boston: South End Press.
- Chomsky, Noam (2000) *New horizons in the study of language and mind*. Cambridge, UK: Cambridge University Press.
- Chomsky, Noam (2013) Notes on denotation and denoting. In: Ivano Caponigro and Carlo Cecchetto (eds.) *From grammar to meaning*, 38–46. Cambridge, UK: Cambridge University Press.
- Chomsky, Noam (2015) Problem of projection: Extensions. In: Elisa DiDomenico, Cornelia Hamann, and Simona Matteini (eds.) *Structures, strategies and beyond: Studies in honour of Adriana Belletti*, 3–16. Amsterdam & Philadelphia, PA: John Benjamins.
- Chomsky, Noam (2016) *What kind of creatures are we?* New York: Columbia University Press.
- Chomsky, Noam (2017) The Galilean challenge. *Inference: International Review of Science* 3.1. <https://inference-review.com/article/the-galilean-challenge> [accessed June 2021].

- Chomsky, Noam (2019a) Some puzzling foundational issues: The Reading program. *Catalan Journal of Linguistics. Special Issue, Generative Syntax: Questions, Crossroads, and Challenges*, 2019: 263–285.
- Chomsky (2019b) UCLA lectures. Invited lectures at the Department of Linguistics, University of California, Los Angeles, 29–30 April, 1–2 May 2019. <https://linguistics.ucla.edu/noam-chomsky/> [accessed June 2021]. (manuscript available with some changes and an introduction by Robert Freidin. <https://ling.auf.net/lingbuzz/005485> [accessed June 2021].)
- Chomsky, Noam (2021a) Reflections. In: Nicholas Allott, Terje Lohndal, and Georges Rey (eds.) *A companion to Chomsky*, 583–594. Hoboken, NJ: Wiley Blackwell.
- Chomsky, Noam (2021b) Linguistics then and now: Some personal reflections. *Annual Review of Linguistics* 7: 1–11. <https://www.annualreviews.org/doi/abs/10.1146/annurev-linguistics-081720-111352> [accessed June 2021].
- Chomsky, Noam (to appear) Simplicity and the form of grammar. *Journal of Language Modelling* 9(1).
- Chomsky, Noam (forthcoming) Genuine explanation. In: Giuliano Bocci, Daniele Botteri, Claudia Manetti, and Vincenzo Moscati (eds.) *Issues in comparative morpho-syntax and language acquisition*.
- Chomsky, Noam, Ángel J. Gallego and Dennis Ott (2019) Generative grammar and the faculty of language: Insights, questions, and challenges. *Catalan Journal of Linguistics. Special Issue, Generative Syntax: Questions, Crossroads, and Challenges*, 2019: 229–261.
- Collins, Collins (2019) Thought as syntax. *Ordinary working grammarian: A blog about natural language syntax, fieldwork and life*. Thursday, October 24, 2019. <https://ordinaryworkinggrammarian.blogspot.com/2019/10/thought-as-syntax.html> [accessed June 2021].
- Collins, Chris and Edward Stabler (2016) A formalization of minimalist syntax. *Syntax* 19: 43–78.
- Epstein, Samuel David (1984) Quantifier-PRO and the LF representation of PRO_{arb}. *Linguistic Inquiry* 15: 499–505.
- Evans, Nicholas and Stephen C. Levinson (2009) The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32: 429–492.
- Fong, Sandiway (2021) Some third factor limits on Merge. Ms.: University of Arizona.
- Gallistel, C. R. and Adam Philip King (2009) *Memory and the computational brain: Why cognitive science will transform neuroscience*. Chichester: Wiley-Blackwell.
- Graeber, David and David Wengrow (2021) *The dawn of everything: A new history of humanity*. London: Penguin Random House.
- Heim, Irene and Kratzer, Angelika (1998) *Semantics in generative grammar*. Oxford: Blackwell.
- Hinzen, Wolfram (2017) Universal Grammar and philosophy of mind. In: Ian Roberts (ed.) *The Oxford handbook of Universal Grammar*, 37–60. Oxford: Oxford University Press.
- Huybregts, M. A. C. (2018) On identifying terms. Workshop: Generative Grammar at the Speed of 90, in honor of Noam Chomsky. University of Arizona, 10–11 December 2018.
- Huybregts, M. A. C. (2019) Infinite generation of language unreachable from a stepwise approach. *Frontiers in Psychology* 10: 425. doi: 10.3389/fpsyg.2019.00425. <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00425/full> [accessed June 2021].
- Kemmerer, David (2012) The cross-linguistic prevalence of SOV and SVO word orders

- reflects the sequential and hierarchical representation of action in Broca's area. *Language and Linguistics Compass* 6.1: 50–66. <https://onlinelibrary.wiley.com/doi/abs/10.1002/inc3.322> [accessed June 2021].
- Kitcher, Philip (1973) Fluxions, limits, and infinite littleness: A study of Newton's presentation of the calculus. *Isis* 64: 33–49.
- Lightfoot, David W. (2020) *Born to parse: How children select their languages*. Cambridge, MA: MIT Press.
- Mann, Charles C. (2006) *1491: New revelations of the Americas before Columbus*. New York, NY: Knopf.
- Martin, Roger and Juan Uriagereka (2014) Chains in minimalism. In: Peter Kosta, Steven L. Franks, Teodora Radeva-Bork and Lilla Schürcks (eds.) *Minimalism and beyond: Radicalizing the interfaces*, 169–194. Amsterdam & Philadelphia, PA: John Benjamins.
- Mayr, Ernst (1995) Can SETI succeed? Not likely. <http://www2.hawaii.edu/~pine/mayr.htm> [accessed June 2021]. (Originally appeared in *The Bioastronomy News* 7.3)
- McElree, Brian and Thomas G. Bever (1989) The psychological reality of linguistically defined gaps. *Journal of Psycholinguistic Research* 18: 21–35.
- Mendivil-Giró, José-Luis (2018) Is Universal Grammar ready for retirement? A short review of a longstanding misinterpretation. *Journal of Linguistics* 54: 859–888.
- Minsky, Marvin (1985) Why intelligent aliens will be intelligible. In: Edward Regis (ed.), *Extraterrestrials: Science and intelligence*, 117–128. Cambridge, UK: Cambridge University Press.
- Morin, Edgar and Massimo Piattelli-Palmarini (1974) *L'unité de l'homme: Invariants biologique et universaux culturels*. Paris: Éditions du Seuil.
- Moro, Andrea (2016) *Impossible languages*. Cambridge, MA: MIT Press.
- Pietroski, Paul M. (2005) Meaning before truth. In: Gerhard Preyer and Georg Peter (eds.) *Contextualism in philosophy: Knowledge, meaning, and truth*, 255–300. Oxford: Oxford University Press.
- Pietroski, Paul M. (2018) *Conjoining meanings: Semantics without truth values*. Oxford: Oxford University Press.
- Quine, W. V. (1956) Quantifiers and propositional attitude. *The Journal of Philosophy* 53: 177–187.
- Reuland, Eric (2011) *Anaphora and language design*. Cambridge, MA: MIT Press.
- Reuland, Eric (2014) Reflexives. Oxford handbooks online. <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935345.001.0001/oxfordhb-9780199935345-e-9> [accessed June 201].
- Rizzi, Luigi (1982) *Issues in Italian syntax*. Dordrecht: Foris.
- Roberts, Ian (2019) *Parameter hierarchies and Universal Grammar*. Oxford: Oxford University Press.
- Roberts, Ian and Jeffrey Watumull (2015) Leibnizian linguistics. In: Ángel J. Gallego and Ott, Dennis (eds.) *50 years later: Reflections on Chomsky's Aspects. MIT Working Papers in Linguistics* 77: 211–222. Cambridge, MA: MITWPL.
- Schein, Barry (2017) *And: Conjunction reduction redux*. Cambridge, MA: MIT Press.
- Seely, T. Daniel and Hisatsugu Kitahara (2021) Structure building under MERGE. Poster presented at the 39th West Coast Conference of Formal Linguistics. University of Arizona, 10 April 2021.
- Shi, Rushen, Camille Legrand, and Anna Brandenberger (2020) Toddlers track hierarchical structure dependence. *Language Acquisition* 27: 397–409.
- Totani, Tomonori (2020) Emergence of life in an inflationary universe. *Nature: Scientific*

- Report* 10.1671. <https://doi.org/10.1038/s41598-020-58060-0> [accessed June 2021].
- Turing, A. M. (1950) Computing machinery and intelligence. *Mind* 49: 433–460.
- Weyl, Hermann (1934) *The Open world: Three lectures on the metaphysical implications of science*. New Haven: Yale University Press.
- Williams, Edwin (1978) Across-the-board rule application. *Linguistic Inquiry* 9: 31–43.
- Yang, Charles (2013) Who's afraid of George Kingsley Zipf? Or: Do children and chimps have language? *Significance* 10: 29–34.
- Yang, Charles (2016) *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: MIT Press.