The Miracle Creed and SMT*

Abstract

Einstein once observed that the "true theorist believes that the logically simple is also the real,…a miracle creed, but it is a miracle creed which has been borne out to an amazing extent by the development of science." The Strong Minimalist Thesis (SMT) holds that language too may satisfy the miracle creed at its core. SMT was proposed only as a guideline for research, but increasingly it seems that it may be valid in surprising ways. Recent work, reviewed and extended here, provides some credibility to the strong interpretation of SMT, with many consequences, if tenable.

*To appear: Matteo Greco & Davide Mocci, eds.

The modern scientific revolution adopted Galileo's precept that nature is simple and it is the task of the scientist to show it. In his own words, Nature "does not act by means of many things when it can do so by means of a few." A later version, backed by solid results, was Leibniz's Principle of Optimality. By now substantiation is so strong that the bold conjecture has been adopted as a norm for scientific theory. As Einstein formulated the conclusion, the "true theorist… believes that the logically simple is also the real,… a miracle creed, but it is a miracle creed which has been borne out to an amazing extent by the development of science."[1] It is also an expectation for scientific progress. John Wheeler anticipated that behind it all is a "central idea…so simple, so beautiful, so compelling" that someday, when we grasp it, "we will all say to each other, `Oh, how could it have been otherwise!'."[2]

There is an independent reason to seek maximal simplicity. The goal of serious inquiry is explanation, which deepens as greater simplicity of theory is achieved. That maxim has been the guiding motive for the "generative enterprise" since its

[1] For review of the history and justification, see McDonough (2022).

[2] Wheeler, J.A. (1986). The actual quote is a little different from the usual rendering on the internet: "Can we ever expect to understand existence? Clues we have, and work to do, to make headway on that issue. Surely someday, we can believe, we will grasp the central idea of it all as so simple, so beautiful, so compelling that we will all say to each other, `Oh, how could it have been otherwise! How could we all have been so blind so long!'." His conclusion is that behind it all is information communicated in a participatory observer-bound universe: "It from Bit" in his formula.

inception.[3] The search is reflected in successive versions of UG over the years. Each step sought a simpler formulation of UG with fewer language-specific conditions (LSCs), hence deeper explanation, leading to another bold conjecture, the Strong Minimalist Thesis SMT.[4]

Within the biolinguistic framework, UG must satisfy two primary empirical conditions.[5] One has been called "Plato's problem": How can a particular language be acquired on the basis of available evidence, the problem of Poverty of Stimulus (POS)? The second is "Darwin's problem": How can the faculty of language FL have evolved under the conditions of human evolution?

The POS problem ("explanatory adequacy") has been a driving concern since the origins of the enterprise. It was always known to be severe, but experimental studies of language acquisition, particularly those of the late Lila Gleitman and her associates, have shown that the severity had been seriously underestimated. The problem of evolution of X, of course, can be addressed only to the extent that we have some idea of what X is, FL in this case. The problem becomes feasible to the extent that UG has achieved a simple enough form. That may well mean that it must approach SMT, given the empirical constraints on evolution of language, which appear to be narrow.[6]

The overarching goal of this line of inquiry is to determine to what extent language is similar to other natural entities in satisfying the miracle creed, its fundamental properties resulting from "third factor" laws of nature, in this case principles of computational efficiency, as is natural for a computational system like language.[7] To the extent that this can be achieved, approaching SMT, both

[3] Chomsky, N. (2020).

[4] The concept of simplicity, a guide for inquiry generally (SMT in particular), is an evolving one. Clarity of goals and guiding principles often emerges in the course of pursuing them.

[5] Understanding UG to be the theory of the faculty of language FL, determining a set of generative grammars, each the theory of a particular language. There are other conditions having do with neural coding, but too little is known about the brain for this to be a significant operative factor today.

[6] Berwick, R. and N. Chomsky (2016). More recent work shows that the window for likely evolution of the core properties of language is narrower than what we assumed on the basis of information then available. Genomic studies place separation of the San from other Homo sapiens (the earliest known separation) in the neighborhood of 100-160kya, all sharing the faculty of language, presumably already in place. Gronau et al. (2011); Kim et al. (2014); Maier et al. (2022).

[7] I will assume as a starting point here the version discussed in Chomsky (2021); henceforth GK. For the revisions and updating here, I am particularly indebted to discussions with Riny Huijbregts.

Plato's and Darwin's problem become more feasible, given the narrow empirical constraints on "learnability" and evolvability.[8]

One guideline for inquiry then is the principle S:

[S] The computational structure of language should adhere as closely as possible to SMT

In particular, we hope to show that as few as possible LSCs enter into the computational structure of language; optimally none, apart from setting the computations in operation by providing the primitive elements to which they apply (lexical items, we assume here) with their properties. For reasons that will become clearer below, we restrict attention here primarily to I-language, understood as generation of the CI interface, independent of Externalization to one or another sensory modality (typically speech).

As a simple model, we can think of bee nests forming hexagons. The bees construct tubes, physics turns them into hexagons. To take a more intricate case, deep homology yields outer hair cells, physics turns them into cochlear spirals in mammals.[9] Similarly, some disruption of the neural basis for cognitive systems yielded recursive enumeration, nature then reconstructed the new system as SMT (so we would like to show).

All of this is consistent with how evolution proceeds quite generally. Schematically, we can distinguish three stages. First some disruption takes place: perhaps a mutation, or genetic drift, or gene transfer, or one microorganism swallowing another, generally assumed to be the breakthrough that led to complex cells. The second stage is reconstruction: Nature reconstructs the new entity in the simplest way, observing the miracle creed, paying no attention to how the new entity might be used.[10] The third stage is winnowing: the outcomes that reproduce more effectively prevail. Natural selection.

[8] "Learning" in quotes, because it is not clear that the category can be seriously distinguished from general growth and development, all aspects of which involve various combinations of the three factors that enter into these processes: innate structure, external data, laws of nature.

[9] Lorimer et al. (2015).

[10] Accordingly, what evolution yields might be seriously dysfunctional, particularly over a long period as circumstances impose new demands, even if at each stage of evolution the optimal system is constructed at the reconstruction stage following disruption.

That suggests a reasonable scenario for language evolution. First, some minor rewiring of the brain took place, yielding the new property of recursive enumeration. The next step is reconstruction: Nature devises the simplest and most elegant way to organize the new system: SMT. Basic conceptual atoms and theta theory become available, possibly with more primitive origins. The computational principles are given, part of natural law. The winnowing stage is never reached, possibly because of lack of time, possibly – and more interestingly – because the optimal system is so delicately designed that it's either all or none, which might be the case, so recent work suggests (see GK).

To illustrate the task we face in pursuing this program, consider perhaps the most fundamental property of language, with rich consequences for cognitive science generally: Structure-Dependence -- a property of I-language. Structure-Dependence follows directly from the optimal assumption. Nature selected the simplest computational operation, yielding the "Basic Property" of FL: recursive enumeration of a digital array of structured expressions, each with an interpretation at the CI interface (essentially as a thought, or instructions for constructing a thought) and optionally in some sensory-motor domain. That is the goal we hope to attain as closely as possible for the various components of I-language – which, we can plausibly assume, exist by virtue of the enabling function of SMT (see GK).

In the case of Structure-Dependence, Andrea Moro's work (2016) provides neurological evidence supporting the conclusion, a very difficult achievement, so far unique but a goal to be sought throughout. As highlighted by the title, his cited book deals with a crucial element in explanation, often overlooked: what is excluded? A true explanation answers the question "Why this, *and not that*?"

The same reasoning has been extended to a number of modules of language (GK), but foundational problems remain and the basic assumptions must be sharpened and modified. That's the topic I will address here.

We take I-language to be a system generating thought, a traditional view that seems increasingly well supported. Several categories of thought are relevant to language structure and use. One category is *propositional*: basic theta-

structure.[11] A second is *clausal*: force- and information-related (interrogative, topic, focus,…). The familiar property of duality of semantics.

A familiar quip in the sciences is that the only numbers are 1, 2, 3 and infinity, with 3 doubtful. We can entertain the doubts: the only numbers are 1, 2, and infinity. The only operations and relations that accord with SMT are unary, binary, and unbounded. Unary operations do nothing in I-language,[12] so we can omit them. Binary relations permit the fewest options for carrying a derivation forward, a sufficient reason for keeping to n=2 for admissible relations and for the operations that create them, unless shown otherwise.

The binary operation of External Merge provides theta structures {X,Y}, where one member receives and the other assigns a theta role, an LSC that underlies the Basic Property. That enables parallel operations, which in turn provide the option of structurally identical inscriptions that are repetitions, not copies, a crucial property of language but not of constructed formal systems, which do not make use of this option.

If the propositional core of language keeps to binary theta structures, there will be no n-ary theta structures for n>2. In particular, complex VPs such as double-object constructions will have internal structure, as commonly assumed. The clausal component also keeps to binary structures, but with no theta-marking.

If I-language is basically a thought-generating system, it optimally should observe the principle T:

[T] All relations and structure-building operations (SBO) are thought-related, with semantic properties interpreted at CI.

T stands alongside of S, the miracle creed for the case of language, as a desideratum.

[11] Subject-predicate (SPEC-INFL) and its VP counterpart (object-raising) are not theta structures, but have secondary semantic properties, along with special properties of phase-internal IM. We put these constructions to the side for now, keeping to the core operations, returning to them later with the core system better established.
[12] NB: in I-language. Without I-language conditions and keeping to the minimal lexicon (one member), Merge yields the successor function, and further steps towards arithmetical knowledge.

Relations are binary. The simplest ones are sisterhood and term-of, the domains of EM and IM, respectively. The next simplest relation after these relates Xε{X,Y} to a term of Y: c-command, the output of IM and domain of other relations. We refer to the structure expressing c-command as a c-command(cc-)configuration. Optimally, language keeps to these: the simplest relations, the only ones linked to theta-theory/Merge, thus conforming to S and T.

Consider the copy-repetition distinction. To satisfy S and T, it must be expressed in terms of admissible (thought-related) relations/operations. The only candidate relation is a cc-configuration. Therefore, we take X and Y to be copies if they are structurally identical and in a cc-configuration. Note that this conclusion is forced if we keep to S,T.

Earlier work (e.g., GK) postulated an operation FormCopy FC which establishes the copy relation in a cc-configuration. We can adopt FC for convenience, but it has no further status; it need not be listed among the admissible operations.[13]

We assume that the primary SBO is Merge, with its two subcases EM and IM. Merge should then be defined so as to satisfy S and T.[14]

The only other permissible relation is unbounded set, with the SBO FormSet (FS). We can assume this to be a costless operation available freely for all inquiry, hence almost never mentioned in the study of language or elsewhere. We use it, for example, in constructing the workspace WS and the lexicon LEX.[15]

Assume for convenience that for XεLEX, X becomes accessible to SBOs only after it is inserted into WS. Merge selects its (binary) domain in the manner discussed in GK. FS selects $X_1,\ldots,X_n$ from WS and yields $\{X_1,\ldots,X_n\}$. Binary FS is distinct from Merge in lacking its special theta-related properties. FS with arbitrary *n* appears in language, but in special ways. We will return to it briefly but keep now to Merge.

[13] The copy relation and FC observe the phase-impenetrability condition PIC, but that is not an independent condition. It follows automatically from phase-based derivation, which renders the contents of a phase invisible at the next step.

[14] On definition of Merge, see GK and for more extensive disc ussion, Chomsky et. al. (to appear); henceforth MRG.

[15] Exactly what is stored in LEX, and how, raises interesting questions.

Adhering to the third factor principle of Minimal Search, Merge applies to members of WS but not their terms. EM selects X, then Y, in WS and forms the theta structure {X,Y}.[16] IM selects XεWS and then a term Y of X. The choice of Y is restricted: only certain terms are "eligible." To keep to principle T, the restriction should be in terms of theta structures. A term Y of X is always a member of a theta structure (GK). Our prime concern here is the recipients of theta roles (theta-marked elements): e.g., NP in {NP,VP}, {R,NP}, {P,NP}. R,P and other heads cannot be raised, so the simplest principle would be to restrict eligibility to the terms that are theta-marked. That reflects standard practice, and also undergirds the intuition that IM is "raising," whereas technically it is symmetric merging of X and a term Y of X.

That decision puts aside movement of theta assigners: VP, AP. These operations have different properties both in form (prosody, etc.) and in interpretation, which is more like topicalization, e.g.:

[1](a) J's novels, I will never read
(b) read J's novels, I never will

With that qualification, the problem of eligibility is resolved for the propositional system.

The central question then is this: How much of I-language can be encompassed by Merge/theta theory and simplest relations and operations, conforming to principles S and T?

The standard view has been that there are three kinds of movement: A, A', head (where head-movement means raising of a head to a head position, not just raising of a head). Head-movement is not properly formulable, and can be eliminated (GK). There is strong evidence that A and A'-movement must be segregated, reviewed in GK, which, however, provides only a loose and incomplete description of the distinction and how it is established.

A-A' movement are close respectively to EM-IM with their theta-theoretic correlates, but not quite. Phase-internal raising (SPEC-INFL/object-raising) is

[16] We put aside adjuncts and such elements as modals, auxiliaries, left-periphery sites, etc., perhaps properly analyzed as features of C, v, scattered in various ways in Externalization.

again problematic, sharing some properties of standard A'-movement but called A-movement. Let's continue to put it aside for now, returning to it later. For the core system, there is no need to distinguish between A- and A'-positions, A- and A'-movement. There's just IM and EM, which are ineliminable, satisfying duality.

The next step is to impose segregation. It is established by IM, which carries the derivation from the propositional to the clausal domain. The simplest way to impose segregation is to keep to that property: IM creates an element that has no further interactions with the EM-generated structures that constitute the propositional domain or with operations that apply there. That is in the spirit of the principles S and T.

For ease of exposition, we can think of the element E that is IM-ed to the phase edge as being put in a box, separate from the ongoing derivation D. E must however be accessible to D at later phase levels for interpretation at the interfaces. The GK system maintained the standard assumption that E moves phase-by-phase to SPEC-phase, where it is interpreted at CI (also at SM if Externalization is activated). Now there's no movement and the boxed element E is accessible at the phase level. If E is not accessible at a phase level it's a subjacency violation, the counterpart to "no escape hatch" in the standard version.

The next step is to look more closely at how EM and IM operate in the box theory, comparing this to the standard theory adopted in GK. Let's take one of the simplest cases, a VP structure with a NP in SPEC-VP,[17] keeping to essentials (subscripts for exposition only):

[2] [NP_1 [$_1$ V NP_2]]

Is NP_1 formed by IM or EM? If by IM, then NP_2 deletes under Externalization and the two NPs are interpreted the same way at CI. Thus if NP is *many people*, [2] will be understood as "many people like those very people" and pronounced "many people like". Plainly that is false. It is understood as "there are many people who like many people" (John Doe, for example, likes many people) and

[17] The NP in SPEC-VP can be the external argument EA theta-marked by VP, or in what is called an A'-position in the standard theory. We return directly to the distinction.

pronounced “many people like many people.” Therefore, it appears that it cannot be formed by IM.

That, however, is incorrect. NP_1 in [2] is in SPEC-VP. But IM to SPEC-VP is legitimate, as in what has been called A’-movement, say Topicalization. Thus [2] would be legitimate if it were formed by IM as the first step of topicalization, with SPEC-VP considered to be an A’-position (yielding ultimately “John/many people, Bill likes ~~John/many people~~”), but it is illegitimate if it is formed by IM with SPEC-VP considered to be an A-position, with NP_1 taken to be the external argument EA. How does the derivation (and its interpretation) know?

In the box theory, IM of NP_2 imposes segregation of NP_1; it is boxed and immune to the processes that apply to EM-generated structures in the system of propositional interpretation. In particular, it is immune to theta-marking. Hence after IM raising of NP_1, VP still has an unassigned theta role, which means that EA must be introduced (necessarily by EM) to be theta-marked by VP. IM-generated boxed NP_1 will be accessed for instructions at later phases. We now have an explanation for the interpretation of [2] at both interfaces, one that extends without comment to further expansion of [2] as in [3]:[18]

[3] NP_1 tried [$_1$ NP_2 to win]]

These conclusions adhere closely to principles S and T. They also have strong empirical support, to which we turn directly.

In the standard theory (taken over in GK), these issues were addressed by invoking a theta-theoretic principle, Univocality:

[U](a) A theta-assigner assigns no more than a single theta-role to an inscription or a copy pair
(b) A theta-position cannot receive more than one theta-role.

The principle U, however, leaves the problems unresolved.

There are, furthermore, problems with invoking U. One is that U(a) is not relevant: the subject theta role is not assigned by V but by V-complement, by the

[18] For independent reasons, IM does not apply here. See note 28.

phrase [$_1$…] in [2]. That is clear for VP idioms and for weak verbs like *take* or *have*, but the same is true more generally as careful analysis of semantic interpretation reveals.[19] U is also complex, invoking copy pairs for (a) but a theta-position for (b) (with no clear application).

There are other issues at a more abstract level. Apart from its unwanted complexity, Univocality is not an essential property of UG. If it is dropped, the rest of the system remains intact. Furthermore, it is an LSC, the kind of assumption we are trying to eliminate in the effort to approach SMT. It also introduces Theta-theory into the computation, which should be purely computational in accord with a strict interpretation of principle S.[20]

In the box theory, all of these problems disappear.

Let's turn to direct empirical confirmation. One major source is the phenomenon of reconstruction, illustrated by such cases as [4], using the conventional notions trace and PRO:

[4](a) one interpreter each seemed t to be assigned to the diplomats
(b) *one interpreter each tried PRO to be assigned to the diplomats

In (a) the matrix subject is raised by IM from t and reconstructed in that position. In (b) the matrix subject is inserted in-situ by EM and there is no reconstruction; rather interpretation of an independent element PRO. Hence *, since *each* is stranded.[21]

We can now eliminate trace and PRO. The distinction in [4] follows directly from the box theory. Both *t* and PRO are copies in a cc-configuration, but only the former, generated by IM with its antecedent boxed, is a reconstruction site.[22] The reason is straightforward. The boxed element lacks a theta role so interpretation must be at the trace position; reconstruction. Not so for PRO, whose antecedent

[19] The issue was raised in the 18th century critique of the theory of ideas, by César Chesneau Dumarsais, then by Thomas Reid, both of whom observed that a phrase like "I have a thought" is distinct from "I have a diamond," and should be understood roughly as "I think." Such phrases are what Gilbert Ryle called "systematically misleading expressions." For details, see Chomsky (1966).
[20] There are also some technical problems, discussed in GK.
[21] For other cases with similar properties, see GK/MRG.
[22] There is no need to check the history of derivation, thus violating the general Markovian property of computation.

has a theta role so there is no reconstruction. The analysis reduces to EM/IM and principle T.

To take a more complex case, consider [5]:

[5](a) someone stole many books from the library
(b) many books were stolen t from the library
(c) many books are easy [to steal t from the library]

Both (a) and (b) mean that there's a thief on the loose. In (b), *t* marks the copy in the reconstruction position left by IM, indicating that interpretation is the same in (a) and (b) for the reasons just outlined.

Example (c) has a different meaning: some category of books, maybe history books, are easy to steal from the library. *t* again is a copy left by IM, but not IM of *many books*, as is clear from the interpretation (differing from the raising case (b)). As we know, the bracketed phrase has the properties of long-distance A'-movement, now revised with a boxed empty element E interpreted at SPEC-C (that is, at []. E is interpreted as an empty operator leaving an open sentence, which functions as a predicate, theta-marking the EM-inserted subject, which is in the same position as the predicate-internal subject in "John saw Bill". The box theory again explains why (c) is interpreted differently from (a), (b).

Suppose that X is IM-ed to Y. It can enter into a copy relation with elements of Y in a cc-configuration (including the position from which it raised). That's the basis for analysis of sequences and ATB in GK. That remains unchanged in the box system. But in the standard (GK) system, if X moves again to a higher phase, it should also enter into copy relations with elements of Y. Not in the box system, however, because X doesn't raise.[23] Consider again simple topicalization [6]:

[6] Bill, John met yesterday

Stripping down to essentials, in the standard version the structure is [7] while in the box theory it is [8]:

[7] {$Bill_3$, {John, {$_{v*P}$ $Bill_2$, {met $Bill_1$ yesterday}}}}

[23] Relevant examples are discussed in GK (21-24), but not dealing with deletion under Externalization.

[8] {C, {John, {$_{v*P}$ Bill$_2$, {met Bill$_1$ yesterday}}}}

In both cases, the copy relation <Bill$_2$, Bill$_1$> holds in a cc-configuration, and under Externalization *Bill$_1$* deletes by the standard economy-based principle.

Suppose that we replace *John* by *Bill*, structurally identical with the topicalized element *Bill*. We then derive [7'] and [8'], respectively, highlighting the elements involved in topicalization:

[7'] {**Bill$_3$**, {Bill$_4$, {$_{v*P}$ **Bill$_2$**, {met **Bill$_1$** yesterday}}}}
[8'] {C, {Bill$_4$, {$_{v*P}$ **Bill$_2$**, {met **Bill$_1$** yesterday}}}}

In [7'] the copy relation <Bill$_3$, Bill$_4$> holds in a cc-configuration, and under Externalization, *Bill$_4$* deletes. Mistakenly. The output is [9] for interpretation:

[9] Bill, Bill' met ~~Bill~~ yesterday

with *Bill'* taken to be a repetition at CI, not a copy (and with the copy of *Bill* deleted under Externalization). The same with such sentences as [8], with *[most people]'* understood as a repetition:

[10] most people, [most people]' regard ~~[most people]~~ as foolish

These facts are unaccounted for in the standard theory (and GK).[24] Now they are. Copies can be formed by IM, but not if IM were to reapply. More generally, what was formerly successive-cyclic wh-movement does not yield copy pairs, another reason to suppose that it does not exist.

In the GK system there were unsolved questions about double object constructions and the copy relation, which arise here too. One way to deal with them would be to assume an applicative PP and adopt Henk van Riemsdijk's proposal about PP being a phase.[25]

[24] Thanks to Emilio Gonzalez for pointing out this problem in the standard theory (GK). Note that some such expressions might be (weak) superiority violations, though not if D-linked. Prosody is also involved: "who persuaded who to leave" vs. *"who perSUAded who to leave" (focal stress on the second syllable of *persuaded*, and destressing on what follows), and many other cases that raise questions whether superiority exists and how its effects might be related to focus.
[25] Van Riemsdijk (1978).

Turning to interpretation, at each phase PH the phase head accesses the box for "instructions". PH is interpreted at CI and (optionally) at SM, the latter typically at the matrix (criterial) position.[26]

Consider some further cases, increasingly complex, all grammatical. Let W = "which pictures of each other," with boldface **W** for the occurrence of W in which *each other* is interpreted at CI. $[_1$ is the phase containing the antecedent of *each other*; $[_3$ is the configuration for anaphora; $[_2$ is the lowest phase, in which W is interpreted as object of *like.*

Consider the standard picture in terms of successive-cyclic wh-movement before raising to SPEC-INFL (avoiding inessentials):

[11](a) W $[_1$ C INFL $[_3$ the-men $[_2$ **W** like W]]]

which pictures of each other do the men like ~~wh- pictures of each other~~

(b) W $[_1$ C INFL $[_3$ the men [**W** hope Bill $[_2$ W like W]]]]

which pictures of each other do the men hope Bill likes ~~wh- pictures of each other~~

(c) W [C INFL [Mary thought $[_1$ C INFL $[_3$ the men [**W** hope Bill $[_2$ W like W]]]]]]

which pictures of each other did Mary think the men hope Bill likes ~~wh- pictures of each other~~

As the examples illustrate, Externalization, anaphora, and thematic relations are dissociated. In (c), Externalization is determined at the matrix; anaphora is determined at the intermediate phase $[_1$ in the configuration $[_3$; and the verb-object relation is determined at the lowest phase $[_2$.

[26] For full overt-movement languages, Spell-out of W is at the matrix position, the "criterial position" in Luigi Rizzi's sense. There can be non-trivial effects at intermediate positions, as discussed extensively in the subjacency literature. See also [12], below.

In the box theory, W only appears in $[_2$. At each phase, W is consulted to determine the next step, providing the instructions just indicated. In (c), for example, the verb-object relation is determined at $[_2$; anaphora is determined at phase $[_1$ in the configuration $[_3$; and Externalization is determined at the matrix. Note that agreement and anaphora keep to the usual cc-configuration, with anaphora penetrating the boxed W. We therefore conclude that what is boxed is inaccessible, but not its terms.

This seems about as simple as possible. Nothing new is added to conventional assumptions; the standard operation of accessibility to W is reformulated. Such cases as [9b,c] were not considered within the standard framework. It appears that mechanisms would have to be more complex.[27]

Let's return to SPEC-INFL (similarly its VP analogue), where NP has raised to SPEC by IM, internally to the phase (A-movement, in standard terminology). There appears to be no need for successive-cyclic A-movement (GK). In that respect, SPEC-INFL is similar to boxed *wh*-. We can extend the generalization by banning IM of SPEC. That follows from restriction of eligibility to theta-marking in the narrow sense, excluding the secondary semantic roles. Nothing is lost. There's no need for raising of SPEC-INFL; anything that looks like that could actually be raising of what occupies its source position. If Externalization is activated, both occurrences will be deleted in a cc-configuration.

Note that nothing has to be said about deletion. It seems reasonable to assume that there is none within I-language. The familiar economy-motivated deletion operation applies if Externalization is activated. It applies to the lower element in a cc-configuration, however it was generated.[28]

Since SPEC-INFL is not at the phase level, it is not accessed at later phases. Therefore, it need not to be boxed. And it shouldn't be, since unlike boxed

[27] Consider, e.g., the problem of dealing with the fact that the interpretation of the lowest phase of the derivation is that Bill likes the pictures of the men, although the antecedent-anaphor relation is determined at a later phase.

[28] If FC applies obligatorily (as is natural, given that it simply identifies a structural relation), it blocks externalization of the lower element of the cc-configuration in such single phase expressions as "many people tried many people to win" or "$John_1$ tried $John_2$ to win." There are ways to unify the two cases (taking *John* to be *Det-John* with a quantificational interpretation), raising issues that would go too far afield. There are unclear factual questions as to whether there is a CI interpretation under distinct reference (Case aside), and no obvious way to answer them.

elements, SPEC-INFL is accessible to other operations (Agree, Labelling, Anaphora).[29]

SPEC-INFL is therefore an anomalous position, similar in some but not all respects to the output of movement to the lowest phase level by IM. It seems that raising to SPEC-INFL is an ancillary operation – which is plausible. The structure is an outlier in the system, related to EPP, Labelling, and secondary semantic roles. The VP counterpart also has anomalous behavior: there are problems about optionality that don't arise in the subject case, also about interpretation (de re, de dicto).

In the box system, an anomaly about the Vacuous Movement Hypothesis disappears. The standard theory adopted in GK requires wh-movement to the criterial position, though by VMH that is not the case for NP-movement (A-movement). Now the requirement is nullified. There may also be a handle on Luigi Rizzi's observations about SM effects lacking at the first raising operation. Perhaps, as Riny Huijbregts suggests, at the *first and only* application of IM raising of wh-XP to SPEC-v*P, the operation accessing the box is vacuous in that it has just been implemented, while at later phases accessing the boxed element is not vacuous in this sense.

Another problem with successive-cyclic movement overcome in the box system, pointed out by Riny Huijbregts, has to do with partial wh-movement, studied by Dana McDaniel (1989). With successive-cyclic movement, in such German expressions as [12] there is a labelling violation at $[_2$, which is not in the criterial position $[_1$ (*t* the variable copy of *wh-XP*):

[12] $[_1$ was [C…$[_2$ wh-XP [C…t…]]]]

The problem doesn't arise in the box theory. At $[_2$, the boxed *wh-XP* is accessed and under Externalization, spelled out, but with no labelling problem since the phrase does not appear in the derivation. At $[_1$, scope of *wh-XP* is determined for

[29] With restrictions. Anaphors do not appear in clausal subject position in the dialects considered here. That bars "they think that each other are in the room" but allows "they think that pictures of each other are in the room." That violates PIC, but anaphora is not generally restricted by PIC: compare "many people like themselves", *"many people like ~~themselves~~" (a cc-configuration, barred by PIC). There is much more to say about this matter of course.

CI, and *was* is spelled out under Externalization. Other questions about the construction remain unanswered, but some at least disappear.

Consider a sample derivation, say, of "who you think that John saw". The first step is IM of object of *see* at the v*P phase:

[13] wh_1 [see wh_2]

wh_1 goes into the box. Its copy *wh_2* is interpreted as direct object. The lower copy *wh2* deletes by the universal economy-based rule if Externalization is activated. There is no need for a special marking in the I-language (DEL) for what is to be deleted, as proposed in earlier versions.

That concludes the phase, which is therefore inaccessible to Merge as the derivation proceeds.

The next step is EM of *John* and INFL, followed by IM of *John* to SPEC-INFLP. Then EM of *that* closes the phase. The lower copy of *John* deletes if Externalization is activated. We then go on to the next clause and the criterial position, where the *wh*-phrase is spelled out and the scope of the quantifier is determined.

Compare [13] with [14], where *$John_1$* is theta-marked by the VP:

[14] $John_1$ [see $John_2$]

On the surface, it looks as though the two should behave the same way. They don't. The reason is timing. In [13], *wh-movement* is within the v*P phase, which then disappears from sight in the derivation. In [14], EM of *$John_1$* is in the next phase. Accordingly, the two occurrences of *John* are repetitions, not copies.[30] That contrasts with [15], a single phase, in which the two occurrences are copies and the lower one deletes in the normal way:

[15] $John_1$ [arrived ~~$John_2$~~]

[30] The assumption here is that the verbal phase is v* with its internal arguments, and that EA enters in the next phase. We might instead take the verbal phase to be [EA v*P], as has commonly been assumed. In that case the distinction between [13] and [14] has to be rephrased; simple timing does not suffice. Rather, IM (being completely within v*P) applies first. Then EM of EA. Then PIC, blocking any relation that crosses the PIC barrier (in particular, the copy relation in the cc-configuration).

The same distinctions hold within the v*P phase:

[16](i) most people believe [most people to be honest] (repetitions, two phases)
(ii) most people try [~~most people~~ to be honest] (copies, single phase)

Let's now consider FS, the free operation available for all inquiry. FS applies to WS, forming the set Z = $\{X_1,\ldots,X_n\}$, $X_i \varepsilon$WS by Minimal Search. One option is n=2, but that is not a case of Merge, as noted above. The derivation will crash unless some process introduces Z into the derivation. Two cases are discussed in GK:

[17](a) John lived in Kansas on a farm near a river
(b) John lived in Kansas on a farm and near a river

The two differ in several respects. (a) is a single event; the farm and river are in Kansas. In (b), the three phrases are independent; John could have lived in Kansas, then moved to a farm in Iowa, then to Cambridge near a river. In (a), raising is possible (with constraints, which may or may not be grammatical). In (b), raising is barred by strong matching conditions, yielding the Coordinate Structure Constraint. In both cases order is imposed in Externalization, but contrary to GK, the ordering imposed by such phrases as *respectively* is a discourse property.[31] Without such a phrase there is no ordering relation, e.g., [18]:

[18] Tom, Dick and Harry danced and sang all night

There is no set of conjuncts that expresses the meaning of [18]. Furthermore, the ordering can plainly be determined beyond sentence boundaries, as in [19]:

[19] John and Bill are good friends. They even write joint papers. They live in New York and California, respectively

Within sentence grammar, sets formed by FS are introduced into derivations in the position of single XP's. A second application of FS (replacing FormSequence of

[31] *See* M. Dalrymple and A. Kehler (1995).

GK) determines what is conjoined and distinguishes compound from independent events.[32]

We might ask whether a single XP should be regarded as a singleton set. Thus in "John left," both *John* and *leave* might be understood as the minimal case of a set. There has been little inquiry into how this might work, among the many issues to be looked at more carefully.

Discussion here and in the work that is presupposed is limited to I-language, with only a few side comments on Externalization. The first step in proceeding beyond I-language would be to consider the transition from I-language to Externalization. That operation might be expected to conform to the guiding principle T of I-language. In important ways, it does.

The central property of Externalization, whatever the SM modality, is linearization (only partial for sign), a condition imposed (we are assuming here) by the non-linguistic SM systems. Linearization observes T in that the non-theta clausal system is external to the theta-based propositional system. Furthermore, it is typically to the left: there is a left-periphery but no right-periphery.

What about the anomalous case SPEC-INFL and its VP correlate, which share some properties with the clausal system and some with the propositional system? With regard to linearization, this case falls together with the clausal system: SPEC is to the left. Note that there is no hierarchical reason for this. The {SPEC, INFLP} structure is symmetric: SPEC c-commands the terms of INFLP, and INFLP c-commands the terms of SPEC.

A broad linearization principle, then, could be that all the non-theta elements, not just those of the clause, are to the left, possibly for pragmatic reasons, which can be relevant to Externalization.

At this very general level, the next question is linearization of head-complement, a mismatch between CI (with no ordering) and SM (which requires it). At the CI-level, V-Object and Object-V have the same interpretation; at the SM level they have to be distinguished. The same for other heads. There seems to be no

[32] See GK for details, now replacing sequences by sets and eliminating the operation FormSequence, the one operation assumed in GK that departed from strict adherence to SMT.

general principle, at least within the domain of this discussion, that provides an answer.[33] The same is true for the fine detail of linearization as in the Rizzi-Cinque cartographic inquiries and much else.

These remarks have been devoted to improving and refining GK/MRG and replacing some basic assumptions of standard versions that GK/MRG incorporated. Needless to say, there is a vast terrain to explore to determine how much of language can be incorporated within this extremely parsimonious framework, approaching SMT. It seems however that we can perceive at least the glimmer of a central idea "so simple, so beautiful, so compelling" that someday, when we grasp it, "we will all say to each other, `Oh, how could it have been otherwise!'."

References:

Berwick, R. and N. Chomsky. 2016. *Why Only Us? Evolution and Language*. MIT press

Chomsky, N. 1966. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought.* 2009. 3d edition. Cambridge U. press. Edited and with an introductory essay by James McGilvray and with improved translations.

Chomsky, N. 2020. "Simplicity and the Form of Grammars." Journal of Language Modelling 9.1: 5-15

Chomsky, N. 2021. "Minimalism: Where are we now, and where can we hope to go." Gengo Kenkyu 160, 1-41 (GK)

Chomsky, N., T.D. Seely, R. Berwick, S. Fong, M.A.C. Huybregts, H. Kitahara, A. McInnerney, Y. Sugimoto. To appear. *Merge and the Strong Minimalist*

[33] In many publications, Richard Kayne has provided ingenious empirical arguments in support of a universal V-O order, an approach that requires many applications of IM that would not fall within the options allowed here in order to yield the observed order. If correct, it would seem to be the only case violating the general conclusion that ordering is imposed by SM conditions and is missing from I-language and CI-interpretation, where it plays no role.

Thesis (Elements in Generative Syntax). Cambridge: Cambridge University Press. (MRG)

Dalrymple, M. and A. Kehler. 1995. "On the Constraints imposed by *respectively.*" https://drive.google.com/file/d/11v0r-wybpSCJx080PCvr_tfnfhRU7Zqj/view

Greco, M. and D. Mocci. To appear. A Cartesian dream: a geometrical account of syntax: In honor of Andrea Moro

Gronau, I., M.J. Hubisz, B. Gulko, and A. Siepel. 2011. "Bayesian inference of ancient human demography from individual genome sequences". Nature genetics. Oct. 43(10):
1031–103

Kim, H. L., A. Ratan, G. H. Perry, A. Montenegro, W. Miller, and S.C. Schuster. 2014. "Khoisan hunter-gatherers have been the largest population throughout most of modern-human demographic history". Nature Communications. Dec.

Lorimer, T., F. Gomez, and R. Stoop. 2015. "Mammalian cochlea as a physics guided evolution-optimized hearing sensor." Scientific Reports 5: 12492. Doi:10.1038/srep12492

Maier, P.A., Runfeldt, G., Estes, R.J. *et al.* African mitochondrial haplogroup L7: a 100,000-year-old maternal human lineage discovered through reassessment and new sequencing. *Sci Rep* 12, 10747 (2022). https://doi.org/10.1038/s41598-022-13856-0

McDaniel, D. 1989. "Partial and multiple wh-movement". Natural Language & Linguistic Theory 7.4: 565-604

McDonough, J.K. 2022. "Leibniz and the Miracle Creed Behind Modern Physics: The Power of Leibniz's Principle of Optimality". https://iai.tv/articles/leibniz-and-the-miracle-creed-behind-modern-physics-auid-2047?_auid=2020&fbclid=IwAR1UxgCghzEtAShZlZ9gwQTC7NgAFpMSKSETjhoQLcWe7qwPEDaWIjgZ75I

Moro, A. 2016. *Impossible Languages*. MIT press

Van Riemsdijk, H. 1978. *A Case Study in Syntactic Markedness*. Peter de Ridder Press

Wheeler, J.A. 1989. "Information, Physics, Quantum: The search for links." Proceedings of the 3rd International Symposium, Foundations of Quantum Mechanics, Tokyo: 354-368. https://philpapers.org/archive/WHEIPQ.pdf