biolinguistics

PsychOpen GOLD

Articles

Check for updates

Creative Minds Like Ours? Large Language Models and the Creative Aspect of Language Use

Vincent J. Carchidi [1]

[1] *Independent Researcher, Philadelphia, PA, USA.*

Biolinguistics, 2024, Vol. 18, Article e13507, https://doi.org/10.5964/bioling.13507

Received: 2023-12-18 • **Accepted:** 2024-09-10 • **Published (VoR):** 2024-10-29

Handling Editor: Kleanthes K. Grohmann, University of Cyprus, Nicosia, Cyprus

Corresponding Author: Vincent J. Carchidi, 409 S. 9th St. Philadelphia, PA, 19147, USA. E-mail: carchidi.vince@gmail.com

Abstract

Descartes famously constructed a language test to determine the existence of other minds. The test made critical observations about how humans *use* language that purportedly distinguishes them from animals and machines. These observations were carried into the generative (and later biolinguistic) enterprise under what Chomsky in his Cartesian Linguistics, terms the "creative aspect of language use" (CALU). CALU refers to the stimulus-free, unbounded, yet appropriate use of language—a tripartite depiction whose function in biolinguistics is to highlight a species-specific form of intellectual freedom. This paper argues that CALU provides a set of facts that have significant downstream effects on explanatory theory-construction. These include the internalist orientation of linguistics, the invocation of a competence-performance distinction, and the postulation of a generative language faculty that makes possible—but does not explain—CALU. It contrasts the biolinguistic approach to CALU with the recent wave of enthusiasm for the use of Transformer-based Large Language Models (LLMs) as tools, models, or theories of human language, arguing that such uses neglect these fundamental insights to their detriment. It argues that, in the absence of replication, identification, or accounting of CALU, LLMs do not match the explanatory depth of the biolinguistic framework, thereby limiting their theoretical usefulness.

Keywords

Cartesian linguistics, computational modeling, creative aspect of language use, generative linguistics, large language models

1 The Cartesian Problem Reformulated

Generative linguists like Noam Chomsky often articulate the Cartesian problem of stimulus-free, unbounded, yet appropriate language use in retellings of linguistic history, a tripartite depiction of human behavior that purportedly demonstrates the exercise of unique intellectual freedom. So central is this "creative aspect of language use" (CALU) to human nature that the "Galilean challenge"[1] is to explain it. Yet, CALU was neglected for much of the modern era, as no intellectual tools existed to deal with the problem of how a finite object—the brain—could yield an infinite array of structured expressions (Chomsky, 2017, pp. 1–2; Chomsky, 2009a, pp. 63–65).

It was not until the work of Alan Turing, Alonzo Church, and Emil Post, among others, in establishing the general theory of computability that one could demonstrate the infinite generativity of a finite system and therefore enable linguists to "address part of the Galilean challenge directly" (Chomsky, 2017, p. 2; see also, Tomalin, 2006, ch. 6). Crucially, the Cartesian problem was *reformulated*—effectively split—for the sake of scientific tractability. Whereas before CALU was a single problem, the generative (and later biolinguistic) approach leveraged the tools of computability theory not to explain language *use* but rather to characterize the computational system that *makes possible* creative language use; a "neurobiological Turing machine" (Watumull & Chomsky, 2020, p. 4). The study of the biological mechanisms *enabling* CALU became possible; how these mechanisms are *used* "remains a mystery" (Chomsky, 2017, p. 2).

This paper argues that the current promotion of Transformer-based Large Language Models (LLMs) as models or theories of human language—typically framed in opposition to the generative or biolinguistic approach—neglects the insights of this history to its detriment. Recent work argues, among other things, that the grammatical output of LLMs is of theoretical significance for linguistics including as a usage-based response to the poverty of stimulus argument (Contreras Kallens et al., 2023), that LLMs acquire knowledge of the abstract rules and statistical regularities of language (Mahowald et al., 2024), and that LLMs are theories of human language that refute the approach most closely associated with Chomsky (Piantadosi, 2023).

The structure of this paper breaks down the biolinguistic approach to CALU through direct engagement with these works. Section 2 explicates CALU from the Cartesian observations and its reformulation in the biolinguistic framework. Section 3 then provides an overview of the status of LLMs in linguistic theory. Section 4 argues that LLMs do not replicate CALU and provide no countervailing reason to invoke an *internalist* orientation and *competence-performance* distinction in the study of human language. Section 5, using this analysis, responds to the argument that LLMs are theories, illustrating how the split

1) The name is drawn from a man so impressed with language—perhaps written communication especially—as "surpassing all stupendous inventions" (Galileo, 2001/1632, p. 120).

of the Cartesian problem in biolinguistics reflects a deliberate *lowering of expectations* for the sake of scientific tractability—a theoretical accommodation LLMs do not permit. Finally, Section 6 links the foregoing analysis to the question of cognitive architecture, arguing that the postulation of a *generative language faculty* that makes possible—but does not explain—CALU offers a more viable account of human language than LLMs permit or indicate.

2 The Creative Aspect of Language Use

2.1 The Cartesian Observations

Chomsky's (1966, 2009a) most extensive treatment of CALU is in his *Cartesian Linguistics*. In it, he traces the observations bound up in CALU to Descartes who argued that a feature of human behavior distinguishing them from non-human animals and machines is how the former *use* their language. This distinction has roots in what Descartes proposed as an *exception* to the "mechanical philosophy"—the idea that all natural phenomena could be explained in reference to their constituent parts and physical contact (Cohen, 1985, pp. 153–154). Language use escapes, Descartes argues, mechanical explanation, making it a useful test for the existence of a "mind like ours" (Chomsky, 1988, p. 5). In his *Discourse on Method*, Descartes lays out this "language test"[2] to be applied to subjects that appear human:

> Of these the first is that they could never use words or other signs arranged in such a manner *as is competent to us* in order to *declare our thoughts to others*: for we may easily conceive a machine to be so constructed that it emits vocables, and even that it emits some correspondent to the action upon it of external objects which cause a change in its organs…but not that it should arrange them variously so as appositely to *reply to what is said in its presence*, as men of the *lowest grade of intellect can do* (Descartes, 1910/1637, p. 60, emphases added).

Descartes' test proposes that a machine possesses a mind like ours if it: utters words in an *intelligible* manner ("competent to us"); expresses a *thought* through variously arranged words of its own *volition* ("to declare" them); and expresses such thoughts through words in a manner that is *appropriate* to the remarks it has heard from others ("appositely to reply to what is said in its presence").[3] This behavior is exhibited by humans of the "lowest grade of intellect"—it is a common human ability.

2) Descartes proposes two tests (see Gunderson, 1964, pp. 197–201), though the second does not concern us here.

https://doi.org/10.5964/bioling.13507

Cordemoy (1668, pp. 4–6) extends Descartes' formulation of the language test, arguing that the distinction between humans and mechanical beings lies in the relationship between words and thought; specifically, the "*facilness of pronouncing Words*" (Cordemoy, 1668, p. 13) is reflective of *thought* (Cordemoy, 1668, pp. 9–13). The key, then, is not the mere production of words but *how* words and other signs are used:

> But yet, when I shall see, that those Bodies shall make signes, that shall have *no respect at all to the state they are in*, nor to their conversation: when I shall see, that those signs shall agree with those which I shall have made to express my *thoughts*: When I shall see, that they shall give me *Idea's*, *I had not before*, and which shall relate to the thing, I had already in my mind: Lastly, when I shall see a *great sequel between their signes and mine*, I shall not be reasonable, If I believe not, that they are such, as I am (Cordemoy, 1668, pp. 18–19, emphases added).

It is not unreasonable to believe that a machine possesses a mind like our own if it makes signs that do not arise in *necessary connection* with its environment; it makes signs that *complement* and *correspond* with the signs that the human interlocutor uses to express their thoughts; it communicates a *novel idea* that nonetheless corresponds with the human's existing thoughts; and the signs *build upon* those of the human.

2.2 The Cartesian Observations in Biolinguistics

Chomsky's approach to the Cartesian observations is to incorporate them in a biolinguistic framework in which human creativity, evidenced most strikingly in ordinary language use, is both enabled and constrained by rule-based properties of the mind conceived as an organic phenomenon (Smith, 2004, pp. 184–185). CALU's role in the biolinguistic framework is thus different than its Cartesian function: rather than a test for the existence of other minds, CALU is an exercise of species-specific *intellectual freedom* (den Ouden, 1975, p. 15; see also, Land, 1974, pp. 16–17).[4] From the Cartesian writings, three features of ordinary human language use are extracted:

2.2.1 Stimulus-Freedom

Linguistic productions are *elicited* by local stimuli, but not *caused* by them. The difference is subtle yet crucial, relating to the distinction between human language use and "purely functional and stimulus-bound animal communication systems" (Chomsky, 2009a, p. 63).

3) The use of "words or other signs arranged in such a manner as is competent to us" reflects the *syntax* of human language; the unique combinatorics that enables the infinite combination of finite elements into new meanings, described in Section 2.2.

4) Note that this does not rule out CALU as a test of other minds within the biolinguistics framework.

Animal communication is restricted to local contexts (Anderson, 2008, pp. 796–797). In contrast, human language use is stimulus-*free*. Linguistic production is not casually tied to a set of stimuli that trigger its uses—no "fixed association of utterances to external stimuli or physiological states" (Chomsky, 2009a, p. 60)—and attempts to trace an expression to a factor in one's environment is merely an "interpretation of an event as part of a pattern…not the well-defined causality of serious theory" (McGilvray, 2001, p. 7). Indeed, in his critique of Skinner's (1957) notion of *stimulus control*, Chomsky wrote that "[w]e cannot predict verbal behavior in terms of the stimuli in the speaker's environment, since we do not know what the current stimuli are until he responds" (Chomsky, 1959, p. 32).

The central problem with the idea of stimulus control is that it stretches the notion of *stimulus* so far as to make it an empty notion: Chomsky's point in his critique of *Verbal Behavior* was that if each linguistic utterance can be accounted for by appealing to the properties of a physical object in the local environment (e.g., saying "red" or "chair" when seeing a red chair), then "the world *stimulus* has lost all objectivity in this usage. Stimuli are no longer part of the outside physical world; they are driven back into the organism" (Chomsky, 1959, p. 52). So, too, for internal physiological states where the disjuncture with animal communication systems is illustrative: there is no specifiable list of physiological states that correspond with utterances. Humans express *thoughts* at will in contrast to the mere expression of *passions* Descartes attributed to animals (Rosenfield, 1968, p. 15).

2.2.2 Unboundedness

Unboundedness is the human ability to produce "an infinite range of discrete, different messages" (Anderson, 2008, p. 797). In contrast to animal communication systems, the uniqueness of human language's code—syntax—lies in its *combinatorics*; the combination of existing and finite parts to form new meanings (Moro, 2016, pp. 21–23). (Greco et al., 2023 find that syntactic information is crucial to the human brain's language processing, casting doubt on a necessary tension between statistical surface distributions and discrete hierarchical structures.) Natural language is characterized in part by its *infinite productivity* (Huybregts, 2019, p. 2). Indeed, Spelke (2010) argues that whereas humans and some non-human animals alike (e.g., rats) possess core knowledge systems, natural language "may serve as a medium for constructing new concepts once words and expressions are linked to representations from multiple core systems" (Spelke, 2010, p. 208). A defining feature of human language is thus its "intricate regulation of form/meaning pairs" (Murphy & Leivada, 2022, p. 1). It does not appear that any non-human communication system exhibits the unbounded generation of hierarchical structures and meaning derived from said structures (i.e., compositionality).[5]

5) e.g., Beckers et al. (2024) on why songbird vocalizations do not show evidence of syntax.

2.2.3 Appropriateness and Coherence to Circumstance

Language use is routinely appropriate to the circumstances of its use and coherent to others who hear (or read) the remarks. This is an obvious statement, but exactly what constitutes the appropriateness condition is elusive. Because language use is stimulus-free, being "appropriate" cannot mean "being caused by environmental conditions" (McGilvray, 2001, p. 9). This rules out a *functional* definition of appropriateness in which language use is strictly oriented towards a specified goal (McGilvray, 2001, pp. 8–10). Language use instead is bound up in *expressions of thought*. Chomsky observes that language use "is recognized as appropriate by other participants in the discourse situation who might have reacted in similar ways and whose thoughts, evoked by this discourse, correspond to those of the speaker" (Chomsky, 1988, p. 5). Thus, rather than *controlled* by external stimuli (making "appropriateness" a matter of causality), language use is connected to external stimuli "by the much more obscure relation of appropriateness" (Chomsky, 1975, p. 302).

Taken together, these three features of language use are "creative." Importantly, they must be present *simultaneously* to achieve this status (Baker, 2007, pp. 236–237). As Collins summarizes: "It is as if the speaker has harnessed randomness (unpredictable unboundedness) in a meaningful, appropriate manner" (Collins, 2021, p. 560).

To be sure, not every sentence in ordinary language use *is* novel; 'stock phrases' are "over-represented in our speech habits" (Dupre, 2024, p. 7). Yet, the essence of CALU is that the ability to produce structured, meaningful expressions is *unbounded*. Linguistic productivity is *infinite*. The class of possible linguistic utterances is unbounded, yet individuals select from this class a meaningful combination of words that is appropriate to their situation but not caused by it (see Collins, 2021, p. 560). It is the *use* of language for purposes beyond the expression of passion, being uncaused yet appropriate, that contrasts with animal communication, which itself reveals its function in humans as a medium of *thought*.

2.3 The Scientific Intractability of CALU

As noted, CALU was split by generative linguists for the sake of *scientific tractability*. CALU appears to fall outside the bounds of *determinacy* and *randomness*, which mark the ends of a spectrum of possible scientific reasoning. CALU is a behavior that is not determined (stimulus-free), can take an undefined form (unbounded), yet it corresponds with the mental states of others who may have no direct knowledge of the speaker's experience (appropriate); it is *neither* determined nor random *yet* it is appropriate. CALU is thus not typically considered amenable to scientific explanation in biolinguistics (see Chomsky, 2009b, p. 200). The implication is that humans are forced to simply "stare in puzzlement at…expression of thought that is coherent and appropriate but uncaused…" (Chomsky, 1995, pp. 38–39), perhaps indefinitely.[6] CALU, taken in its entirety, is *scientifically intractable*.

The reasoning typically employed is premised on scientific inquiry as an exercise of human cognition, with some problems remaining incomprehensible owing to "our peculiar cognitive makeup…it does not indicate any objective profundity or divine design" (McGinn, 1993, p. 87). If the human mind is unable to make sense of this tripartite behavior—unable to offer some account of the phenomenon that deepens our understanding of it—then any science the human mind constructs will be similarly limited. The generative approach remains consistent with this disposition towards science up to the present, in which the starting assumption of any scientific discipline is "that the world can be understood" (Fox & Katzir, 2024, p. 72).

This is a guiding point—generative linguists did *not*, armed with the tools of post-Turing computability theory, aim to resolve the question of why individuals *use* their language in a particular way in any arbitrary instance. Rather, the characterization of a computational system that *enables* the free use of language became the target of inquiry, in this way segmenting the underlying *mechanisms* of language off from the *use* of these mechanisms in ordinary life. This point guides our analysis *because* the scientific intractability of CALU establishes a route to an internalist orientation and a competence-performance distinction in the study of language, among other crucial theoretical maneuvers including the postulation of a generative language faculty. LLM-driven arguments simply neglect CALU, the distinctive challenge it poses to a science of language, and its implications for theory-construction.

2.4 CALU's Theoretical Significance and LLMs

CALU is "the central fact to which any significant linguistic theory must address itself" (Chomsky, 1964, p. 7). I argue this significance is demonstrated in theory-construction in three ways:

> (1) A language *user* is defined in part by the stimulus-freedom, unboundedness, and appropriateness of their language use. The full scope of the problem of ordinary language use is scientifically intractable because it is neither determined nor random yet appropriate. This set of facts therefore justifies an *internalist* orientation in the study of language and the derivation of a *competence-performance* distinction to make part of the problem scientifically tractable.
> (2) CALU's reformulation in the biolinguistic framework reflects a *lowering of expectations* for what a science of language can realistically accomplish, targeting competence (an internal cognitive system) rather than performance because of the latter's scientific

6) Chomsky does not definitively claim that CALU will remain inaccessible to scientific explanation. He suggests that it is a *likely candidate* for the "mystery" category (see Collins, 2021).

> intractability. This complements the "Galilean" method; specifically, it complements the Galilean disposition toward *abstraction* and *idealization* of "the" language faculty and a lowering of scientific expectations.
> (3) By making part of CALU scientifically tractable, the biolinguistic approach lays claim to the cognitive architecture necessary to enable—but not explain—each component of CALU. The postulation of a *generative language faculty* characterizable in computational terms with free access to its resources serves this end.

In contrast, LLMs fail to deal with CALU on each count:

> (1) LLMs are not language *users* in the Cartesian sense. Moreover, they fail to grapple with the full scope of the problem of language use. Recruiting them as models or theories of human language risks neglecting the intractability of CALU. This in turn neglects the need for an internalist orientation and a competence-performance distinction in the study of human language and provides no sufficient alternative way of dealing with the problem of CALU.
> (2) LLMs-as-theories fail to identify, explain, or otherwise shed useful theoretical light on CALU in humans. Failing to offer an independent characterization of CALU, LLMs subsequently fail to justify *not* abstracting away from ordinary language use and lowering scientific expectations. LLMs thus do not match up to the theoretical depth of the biolinguistic approach and complementary Galilean method; they are too ambitious.
> (3) The use of LLMs as models of human language risks failing to make at least part of CALU scientifically tractable. They thus do not enable scholars who leverage them to lay claim to the *cognitive architecture* necessary to enable the uncaused but appropriate expression of thought. This in turn carries the parallel risk of *underestimating* the difficulty of explaining human language and the theoretical accommodations that must be made to this end.

With that, we turn to the status of LLMs in linguistics.

3 The Status of LLMs in Linguistics

3.1 What Is a Large Language Model?

It is worthwhile to clarify what we mean by "Large Language Models." The most prominent types of LLMs are "Generative Pre-Trained Transformers," or GPTs, built from

PsychOpen GOLD

the Transformer architecture (Vaswani et al., 2017). The increasing successes of GPT-1 (Radford et al., 2018), GPT-2 (Radford et al., 2019), and GPT-3 led to an effort to build a system that can generalize beyond its training data by scaling up its internal capacity and training dataset (Brown et al., 2020, pp. 3–10). The effectiveness of this technique compelled researchers to build *dialog agents*—models that can engage in specialized conversational tasks, like Google's LaMDA family of Transformer-based dialog models (Thoppilan et al., 2022). In November 2022, OpenAI (2022) released ChatGPT: a conversational agent underpinned by an LLM (a modified version of GPT-3). ChatGPT-4, a significantly larger model, was released in March 2023 (OpenAI, 2023).

The process of building an LLM begins with the construction of a training dataset consisting primarily of text-based data. This text-data is inputted into the models via "tokens," which represent words or parts of words—the model "reads" words as tokens. During "pretraining," LLMs are given an objective: often, it is to predict the *next* token based on a specific number of previous tokens. The model then attempts to match the predicted token against occurrences in its training data and sends this feedback through the model to update its internal parameters—a process called "backpropagation" (see generally, Mahowald et al., 2024, p. 522). Importantly, LLMs are *statistical* models in that they are models of the "statistical distribution of tokens in the vast public corpus of human-generated text" (Shanahan, 2023, p. 2). Asking an LLM a question is effectively the same as saying, "Here's a fragment of text. Tell me how this fragment might go on. According to your model of the statistics of human language, what words are likely to come next" (Shanahan, 2023, p. 2)?

An LLM is not necessarily a *chatbot.* Moreover, when a user interacts with a chatbot like ChatGPT, they are *not* interacting with the LLM. In conversational agents, the LLM is "embedded in a larger system to manage the turn-taking in the dialogue" (Shanahan, 2023, p. 4). Moreover, the system must also "be coaxed into producing conversation-like behavior. Recall that an LLM simply generates sequences of words that are statistically likely follow-ons from a given prompt" (Shanahan, 2023, p. 4). LLMs are the *base* of a more complex system. The conversational behavior of systems like ChatGPT is the result of a technique for aligning language models' output with human expectations called Reinforcement Learning from Human Feedback (RLHF) (Ouyang et al., 2022).[7]

For our purposes, the following terminology will be employed to make a simple distinction: when referring to LLMs embedded in conversational systems, the term "LLM-powered chatbot" will be used; otherwise, the term "LLM" will be used.

7) RLHF is a complex process, but further details are not warranted.

3.2 LLMs in Linguistic Theory

The use of deep neural networks to intervene in linguistic debates both pre- and post-dates the ChatGPT enthusiasm (e.g., Lakretz et al., 2021; Warstadt et al., 2019; Wilcox et al., 2024). What follows is a survey of more recent research debating the proper role of LLMs in linguistics. The works selected are relevant to our subsequent discussion of CALU and LLMs.

Contreras Kallens et al. (2023) argue that Transformer-based LLMs fulfill the promise of earlier connectionist models: they provide an "existence proof that the ability to produce *grammatical* language can be learned from exposure alone without language-specific computations or representations" (Contreras Kallens et al., 2023, p. 2). On this view, LLMs demonstrate that only domain-general mechanisms are required for human-like language acquisition, contra nativists who posit domain-specific mechanisms. They qualify their view, importantly, to note that they are *not* claiming LLMs are language *users* nor that they *understand* language—they are not users because they lack cognitive capacities supporting social interaction. Their central claim is that LLMs' *grammatical output* is theoretically significant for cognitive science (Contreras Kallens et al., 2023, p. 3). The authors frame LLMs as a "usage-based answer to the [Poverty of Stimulus] argument" in which humans 'memorize, abstract, and generalize' linguistic data in language acquisition and its use (Contreras Kallens et al., 2023, p. 3).

More wide-ranging research by Mahowald et al. (2024) argues for a distinction between *formal* linguistic competence and *functional* linguistic competence—the former refers to "the knowledge of rules and statistical regularities of language" whereas the latter refers to "the ability to use language in real-world situations" (Mahowald et al., 2024, p. 518). They argue that LLMs *do* acquire the abstract rules of human language to a significant extent (e.g., hierarchical structure and abstraction)—thereby serving as viable tools in the study of *human* language acquisition and processing. The authors further argue that LLMs do *not* achieve functional linguistic competence as this is done "in tandem with non-language-specific capacities in real-world circumstances" (Mahowald et al., 2024, p. 519).

Importantly, the formal-functional distinction is justified because "language is robustly dissociated from the rest of high-level cognition" (Mahowald et al., 2024, p. 520), and a 'good at language -> good at thought' and 'bad at thought -> bad at language' fallacy leads the study of LLMs' linguistic (and non-linguistic) abilities astray (Mahowald et al., 2024, pp. 517–519). Thus, *formal* linguistic competence is not sufficient for real-world usage, and modular architectures that integrate language with additional systems are required, much like the human brain/mind (Mahowald et al., 2024, pp. 522–535).

Moreover, Piantadosi (2023) argues "that language models should be treated as bona fide linguistic *theories*" (Piantadosi, 2023, p. 7).[8] These theories "develop representations *of* key structures and dependencies" (Piantadosi, 2023, p. 7) and successfully integrate syntax and (some) semantics "without incorporating any of Chomsky's key methodological claims, like…competence vs. performance, respect "minimality" or "perfection," and avoid relying on the statistical patterns of unanalyzed data" (Piantadosi, 2023, p. 15). Piantadosi (2023, pp. 26–28) also argues that the success of LLMs at capturing human-like linguistic output refutes (in their capacity as theories) the "Galilean method" employed by Chomsky; specifically, they refute the search for underlying principles at the expense of broad data coverage.

A flurry of responses emerged to Piantadosi. Katzir (2023) argues that ChatGPT-4 generates different acceptability judgments about sentences in English, fails to distinguish between linguistic *competence* and *performance*, and lacks a distinction between *likelihood* and *grammaticality*, among other disanalogies that together indicate LLMs are "not suitably biased" to acquire human grammatical knowledge (Katzir, 2023, p. 3). Milway echoes this by arguing that LLMs' learning is *less constrained* than that of humans (Milway, 2023). This point about (un)constrained learning is taken up in greater detail by Kodner et al. (2023) who argue that Piantadosi's use of LLMs to explain human language acquisition neglects the original purpose of the poverty of stimulus argument: namely, "children generalize from their limited input in specific ways, navigating a constrained space of possible natural language grammars" (Kodner et al., 2023, p. 2).

Dupre (2024) argues computational modeling idealizes human language learning *too much*. The theoretical usefulness of LLMs depends on them being sufficiently analogous to human learners. For proponents of computational modeling, the generation of strings is (or assumed to be) generated by an underlying rule or constraint system (i.e., a grammar) (Dupre, 2024, pp. 1–3). But, in generative linguistics, a grammar "describes not the set of public symbols produced or producible by competent speakers, but the laws governing the language-specialized mental faculty underlying such usage" (Dupre, 2024, p. 3). The expression of a *hierarchically structured mental representation* recruits greater cognitive activity than just the language faculty (Dupre, 2024, pp. 3–4)—meaning the adequacy of computational modeling depends on 'factoring out' the myriad causal sources that exist *in addition* to language that comprise a learner's dataset (e.g., communicative goals, externalization systems, etc.).

This highlights the extent to which computational models are perhaps too *idealized*—that a grammar can be learned by these models from data generated *by* the grammar is not tantamount to showing that it can be learned *from* the data of myriad systems (Dupre, 2024, p. 5). The "disunified" process of human language acquisition—contra Piantadosi's singular, data-driven process—indicates the following:

> If there were a plausible way to infer solely from primary linguistic data to a target grammar, why could linguists not simply make

8) The conception of deep neural networks as theories is borrowed from Baroni who argues that it is "appropriate…to look at deep nets as *linguistic theories*, encoding non-trivial structural priors facilitating language acquisition and processing…a general theory defining a space of possible grammars" (Baroni, 2022, p. 7).

> use of such an inference, rather than drawing on data from other languages, experimental research, judgements about unacceptable (and thus unuttered) sentences…and much else (Dupre, 2024, p. 11)?

The implication is that computational models, like LLMs, may not be adequately modeling *human* language acquisition.

Beyond responses to Piantadosi, additional work systematically tests three GPTs' response stability regarding the grammaticality of prompts, finding that all models exhibit "marginal overall above-chance accuracy and absence of response stability" (Dentella et al., 2023, p. 6). These results are interpreted to indicate that the models simultaneously *appear* to master the form of language but do not produce the level of accuracy and stability in grammaticality judgments that *should* result from this mastery (Dentella et al., 2023, pp. 6–8). Hu et al. (2024) responded with re-evaluations, arguing that Dentella et al.'s tasks differed between models and humans, thereby negatively affecting the former's results (Hu et al., 2024, p. 3). Leivada, Dentella, and Günther (2024, pp. 2–3) responded, arguing that in the absence of systematic testing that captures the underlying *reasoning* used in specific tasks by humans and LLMs, an inference from an LLM's human-level performance to its possession of a human-like competence is invalid.

Finally, Moro et al. (2023, p. 84) argue that LLMs can acquire "impossible languages" and "possible languages" with equal facility, a sharp disanalogy with humans who can only deal with the former as a *puzzle* rather than as a grammatical structure (Chomsky & Moro, 2022, pp. 21–22). To be sure, the claim of *equal* facility may be exaggerated, but the disanalogy remains real. Kallini et al. (2024) trained GPT-2 models to test whether they can acquire impossible languages. They find that models trained on possible languages learn more efficiently; the models "prefer natural grammar rules";[9] and models develop human-like solutions to non-human grammatical patterns (Kallini et al., 2024, p. 2). The authors conclude that the models "do not master our set of synthetic impossible languages as well as natural ones…" (Kallini et al., 2024, p. 9). While the authors frame these results favorably for the use of LLMs as models of human language, that GPT-2 models *can* acquire impossible languages, albeit less efficiently than possible ones, *is* a disanalogy with human linguistic competence.

An explicit discussion of CALU is conspicuously absent in this literature (for a brief exception, see Moro et al., 2023, p. 83). Yet, the absence of CALU in the invocation of LLMs in linguistic theory is odd, indicating that a full account of what constitutes *ordinary human language use* is missing. This itself has significant downstream effects on explanatory theory-construction, to which we now turn.

9) They are more surprised by ungrammatical constructions when trained on possible languages.

4 A (Human) Language User Is Creative

Contreras Kallens et al. (2023) conceded that LLMs are not language users owing to their lack of requisite cognitive capacities underpinning social interaction. While this hints at the authors' usage-based leanings in defining what *does* count as a language user, CALU offers a different path: A language user expresses new thoughts and ideas—which may have no roots in their personal history—in a manner that is uncaused by their local context yet appropriate to the situation in which they speak and corresponding with the thoughts of others. A language user is *creative*. Yet, the authors superimpose a usage-based approach onto LLMs, arguing that their "[extrapolation] over past chunks of input" (Contreras Kallens et al., 2023, p. 3) is consistent with usage-based approaches and an answer to the poverty of stimulus argument in the production of grammatical language.

To do this without consideration for CALU is theoretically premature. This turns on the *internalist* orientation and the derivation of a *competence-performance* distinction justified in reference to CALU. LLMs' output, I argue, does not provide a countervailing reason for these methodological moves *because* they are not language users. Let us flesh this out:

4.1 Stimulus-Controlled

LLMs are controlled by a combination of internal and external stimuli in predictable ways; they are not stimulus-free (Moro et al., 2023, p. 83). A Transformer-based LLM is (typically) designed to predict the next token based on a specified number of previous tokens. When provided with an input, the model predictably provides a likely continuation based on the statistical probabilities associated with said tokens. While the model is stochastic (Bender et al., 2021, pp. 616–617), this does not imply stimulus-freedom as its operations are *determined*. Three facts hold steady: (1) The model *will* generate an output value; (2) The output value generated will be a *likely continuation* of the input value; (3) The model will not decide, *absent internal or external instructions to the contrary*, to not generate an output value, to alter the process of token generation, or to initiate its own conversations.

Of course, an LLM can be provided, for example, with internal programming instructions that *do* lead the model to not generate an output based strictly on the most likely continuation of an input. The model can be instructed to sometimes select the second- or third-likeliest next-token. The basic point remains: the model is bound to stimuli.

In LLM-powered chatbots, the situation is fundamentally the same: while the base LLM is not directly queried by end-users, the model responsible for filtering inputs to the base LLM is controlled by the programming instructions set forth by humans while the base LLM is bound by the inputs (external stimuli) it is fed. Stimulus-freedom does

PsychOpen GOLD

not emerge either from the size of the base LLM or the complexity of a conversational system.

LLMs and LLM-powered chatbots are stimulus-controlled in highly specific and predictable ways; they are *impelled* to act in certain ways, but never *inclined*.

4.1.2 Weakly Unbounded

There are two ways to interpret the unboundedness of language, one weak and one strong. The weak interpretation is one in which language use is comprised of an unlimited variety of word combinations, at least according to a structure or configuration (see McGilvray, 1999, pp. 85–86). This interpretation makes no reference to the *subject's* understanding of the *meanings* of these expressions. LLMs plausibly achieve this weak unboundedness given their ability to combine and re-combine words based on the statistical model the LLM builds of human text-data. Indeed, syntactic novelty has been experimentally uncovered in LLMs including GPT-2, with the caveat that one cannot *assume* text is always novel (McCoy et al., 2023).

That said, the Cartesian observations about human language use are not about infinite string-generation. Rather, the problem of CALU is one in which language serves as an uncaused "general instrument of thought and self-expression…" (Chomsky, 2009a, p. 64). This is the strong interpretation. As Section 4.1 shows, the weak interpretation is better suited to the Turing Test.

The problem is that however much it *appears* that LLMs are mastering the form of natural language (*a la* Dentella et al., 2023), that this is tantamount to acquiring a capacity for *thought* and self-expression is entirely unclear. Human language interfaces with other cognitive systems in the generation of structured expressions that convey *meaning*—form/meaning pairs, not string-organization (Murphy & Leivada, 2022, p. 1); it is doubtful LLMs possess a similar linguistic competence. There are at least five reasons that suggest this based on the literature reviewed above: (1) LLMs cannot distinguish between the *likelihood* of an utterance and its *grammaticality* (Katzir, 2023); (2) LLMs' learning is too unconstrained (Milway, 2023) and neglects humans' *poverty of stimulus* dilemma (Kodner et al., 2023). The poverty of stimulus argument, in its most truncated form, refers to a general framing of biological development, carried into the domain of human language, in which developmental outcomes are *underdetermined* by environmental stimuli—that an account of biological development is not adequately grounded solely or even primarily in the *experience* of an organism. This directs a research program toward the organism's innate biological endowment in understanding the outcome, in this case the outcome being human linguistic knowledge (see Berwick et al., 2011, pp. 1207–1210); (3) Computational modeling idealizes *too much* and neglects the actual complexity of human language learning (Dupre, 2024); (4) LLMs lack the stability and accuracy of human grammaticality responses in systematic testing (Dentella et al., 2023); and (5)

LLMs may prefer possible languages but *can* acquire impossible languages (Kallini et al., 2024).

Additionally, LLMs' "hallucinations" (see Ye et al., 2023, p. 2) are mismatched with natural language. The hallucination problem relates directly to the matter of pairing form with meaning—an apparent deficiency in LLMs (see Leivada, Dentella, & Murphy, 2024), and one that highlights Dupre's (2024) point that human language learning is a more complicated affair than computational modeling admits. All this is a reminder that LLMs' human- or above-human-level performances on specific tasks run the risk of being fallaciously interpreted as "unlicensed generalizations" that move from human-like performance to human-like competence, thereby "[reversing] the nature of the argument" (Leivada, Dentella, & Günther, 2024, p. 3).

Neither LLMs nor LLM-powered chatbots are unbounded in the strong sense; at most, they achieve a weak form of unboundedness, but not the relevant kind.

4.1.3 Functionally Appropriate

One of the central achievements of the LLM-powered chatbots is the relative success in aligning their outputs with the expectations of human end-users; generating appropriate responses to inputs and refusing to comply with certain requests deemed illegal or unethical by human programmers. It may seem, then, that these systems—specifically, LLM-powered chatbots that undergo RLHF fine-tuning—meet the appropriateness condition of CALU.

As noted, however, creative language use is not appropriate merely because it is *functional*; "clearly [serving] a need or goal" (McGilvray, 2001, p. 9). RLHF *does*, in contrast, serve a clearly defined goal: to align the outputs of the model with the expectations and intents of a human end-user. The conversational system that results from this process is one whose entire conception of appropriate language use (forgive the anthropomorphizing) is responding to the queries of humans in a manner that is deemed useful, ethical, and legal on *their* terms. This is functional, and thus not appropriate in the Cartesian sense. The purely functional nature of their responses is also exemplified by the fact that they are stimulus-controlled; their "appropriateness" is not *related* to external stimuli but *caused* by them.

Furthermore, base LLMs, disconnected from a conversational system, are not appropriate for a straightforward reason: the need to align their outputs with the expectations of end-users indicates the model cannot use language appropriately.

Taken together, neither LLMs nor LLM-powered chatbots simultaneously exhibit the three features of ordinary language use and thus do not replicate CALU.

4.2 The Turing Test Objection

One objection is that LLMs *are* language users because they pass Turing's (1950) "Imitation Game" in which a human must judge whether their anonymous interlocutor is a human or a machine. Should the machine be judged a human (a sufficient number of times), then it passes the test. Indeed, there is evidence that Turing was influenced by Descartes' language test[10] in the Imitation Game's design (see Abramson, 2011), at which point one might suggest passing Turing's test amounts to passing Descartes' test. Recent experimental results indicate the former is a live possibility (see Jones & Bergen, 2024).

Even if we grant that GPT-4 and comparable LLMs (or LLM-powered chatbots) passed the Turing Test, this is not tantamount to passing the *Cartesian* test. Pulman (2018) argues that the Turing Test is considerably more limited than the Cartesian language test, referencing in part the questioning-answering format typical of the Imitation Game.[11] This format's limitations go deeper, not least of which is that the anonymous machine has *no choice* but to participate in the exchange (i.e., it is stimulus-controlled from start to finish).

McGilvray argues that even if a machine *does* pass Turing's test, "no fact of the matter has been determined; no scientific issues resolved. The test offers no evidence in favor of a specific science of mind, and it does not show that the mind works the way the computer that passes the test does" (McGilvray, 2009, p. 114). Passage of the test simply offers one reason why one *may* begin to start using the language of 'thinking machines' (McGilvray, 2009, p. 114). Chomsky (2009c) observes, tying these strands together, that the steady progression of science following Newton and the collapse of the mechanical philosophy meant that Turing's Imitation Game was formulated within "an entirely different framework" than the Cartesian tests for other minds (Chomsky, 2009c, p. 105). Turing's ambitions were lower than the Cartesians', not aiming at scientific objectivity about possession of a human-like mind (or intelligence).

The lowered expectations of Turing's test bring to light fundamental problems in testing for CALU. The description of stimulus-free, unbounded, yet appropriate behavior is the result of observations made first and foremost of human beings; human behavior is the *baseline*. It is not that the basis for such an ability is, in principle, not replicable on silicon substrates. The problem, in its most abstract formulation, is that any human invention that is routinely subject to the stage-setting familiar of contemporary Turing tests *or* bounded by the direction of a human to produce the outputs in which we search for Cartesian creativity makes such a search nearly self-defeating. It is not self-defeating in principle, as it is possible to imagine a machine that, once developed, acts autonomously in ways that do *not* require the prompting familiar to LLMs or programming instructions and manual patching to keep its unbounded outputs tethered to human

10) Rees (2022) raised the *possibility* that LLMs including GPT-3 and LaMDA crossed Descartes' language threshold—but not in the biolinguistic framework we adopt here.

11) Pulman also argues that the Turing Test effectively tests for *intelligence* whereas the Cartesian probes the possession of a *mind*.

PsychOpen GOLD

reality.[12] But the fact that human understanding does not appear capable of penetrating CALU contributes to the dilemmas in testing for it.[13]

Yet, nothing stated here contradicts our foregoing analysis of LLMs and their chatbots in relation to CALU's three components. Passing the Turing Test, as it is contemporaneously conceived, would not necessarily indicate that they exhibit Cartesian creativity. Our conclusion that LLMs are not language users remains.

4.3 Internalism and the Competence-Performance Distinction

Humans *specifically* are language users. Being clear that a "language user" exhibits CALU is therefore fundamental to a scientific account of human language; this description informs subsequent inquiry. This owes to the importance of identifying a mind capable of species-specific intellectual *freedom*. Yet, neglected in the observation that LLMs are not language users by Contreras Kallens et al. (2023)—and the superimposition of a usage-based approach—is that the characteristics of human language use justify an *internalist* orientation and a *competence-performance* distinction in its study.

The post-Turing split of the Cartesian problem of language use into a matter of *characterizing* a computational system and the *use* of this computational system—with the latter relegated to the domain of near-mystery—should not give the misleading impression that CALU is no longer operative in theory-construction. For Chomsky, and implicit in the foundations of the biolinguistic enterprise, the "creative aspect observations, along with the poverty of stimulus observations, offer a set of facts with which his and—he holds—any science of language must contend" (McGilvray, 2001, p. 5). CALU, recall, refers to language use that is *neither* determined nor random *yet* it is appropriate; uncaused but appropriate expression of thought. The way these facts are dealt with is not to explain actual language use, as this is "a scientifically intractable aspect of the world" (Asoulin, 2013, p. 235). Any theory that attempts to explain uncaused but appropriate language use yields "no scientifically interesting regularities" (Asoulin, 2013, p. 240).

Rather, it is precisely in the formation of an explanatory theory that aims to make *part* of this problem scientifically tractable. Thus, an *internalist* science of language that aims at the mechanisms that make creative language use *possible*—but not the use itself—is justified (Asoulin, 2013, pp. 241–242; see also McGilvray, 2009, p. 2). From this, a distinction between linguistic *competence* (the mechanisms) and linguistic *performance* (the use) is derived. This distinction only makes sense in a scientific framework if one can meaningfully engage with the idea that a finite biological system (i.e., the human brain) can yield infinite outputs; that "[i]t is possible to invent a single machine which

12) For a discussion on the Cartesian test and Chomsky's critique of behaviorism unrelated to LLMs, see Land (1974).

13) I am grateful to Mark Baker for his substantive and clarifying feedback on the relationship between LLMs, the Turing Test, and CALU. The conclusions here are my own.

can be used to compute any computable sequence" (Turing, 1937, p. 241). As a naturalistic inquiry, the biolinguistic approach forgoes Descartes' attribution of the capacity for infinite linguistic generativity to the possession of a "spiritual entity" (see Riskin, 2016, p. 63).

The superimposition of a usage-based approach onto LLMs to signify their theoretical importance is sharply limited if it fails to establish this baseline definition of what counts as a language user. An internalist orientation and a competence-performance distinction are justified in reference to the phenomenon that *does* exhibit CALU—humans—and it is unclear why LLMs' output (however grammatical it appears to humans) should change this. That said, if LLMs *did* reproduce CALU, this would mean that they are the only other (known) phenomenon that is capable of this behavior and it would buttress their use in the study of human language. However, this would not settle the matter of how to conduct a science of human language (see Section 4.1), not least of all because humans are *biological phenomena* whose creative language use may be enabled by a different set of underlying capacities.

5 LLMs-As-Theories Fail to Address CALU

5.1 LLMs-As-Theories and Data Coverage

CALU offers a set of facts that justify an internalist orientation in the study of language and a competence-performance distinction. Implicit in this reasoning is a *lowering of expectations* for what a science of language can accomplish in the face of CALU's intractability. As a result, biolinguists are inclined to abstract away from ordinary language use and idealize "the" language faculty—a disposition embedded in the "Galilean" method. We thus turn to Piantadosi's argument that LLMs are *theories* of human language that serve as a "refutation of the "Galilean" method" (Piantadosi, 2023, p. 27) without reference to methodological moves like the competence-performance distinction.

Piantadosi's argument is this: many natural systems exhibit high-level behaviors that may be surprising if one *only* extrapolates from the underlying rules of the systems—this phenomenon is known as *emergence* (e.g., individual traders in a stock market follow simple rules, yet the market-level effects "are the emergent result of millions of aggregate decisions", Piantadosi, 2023, p. 27). Coupled with the fact that some systems are massively complex, the only viable way to test principles that explain their behavior is through *simulation*—and computational tools are best suited to this task. LLMs are *simulations* of human language, indicating that their grammatical and semantically coherent outputs are emergent results of explanatory value. Such simulations show the futility of the pursuit of *underlying principles* at the expense of data coverage (a pursuit characterized as the Galilean method, Piantadosi, 2023, pp. 26–28).

Piantadosi neglects a critical relationship between CALU and the Galilean method to which we now turn.

5.2 CALU and The Galilean Method

Chomsky's earliest extended invocation of the 'Galilean style' is in his 1980 *Rules and Representations*. It is preceded by Chomsky's (1980, pp. 6–7) articulation of the distinction between "problems" and "mysteries"—those issues where the human intellect can make meaningful progress and those that exceed its grasp, respectively (see Collins, 2021 on this distinction). In tandem, Chomsky articulates the Cartesian view "that we can profitably study motivation, contingencies that guide action, drives and instinct…But…the freedom to choose remains, and remains inexplicable in these (or any) terms" (Chomsky, 1980, p. 8). Chomsky's view is that CALU—which falls into the "freedom to choose"—is closer to a "mystery" than a "problem."

Directly *after* these remarks Chomsky invokes the "Galilean style," citing theoretical physicist Steven Weinberg, deeming the style to be a "narrower version" of the Cartesian thesis that human action cannot be meaningfully studied, but its influences and underlying mechanisms *can* be studied (Chomsky, 1980, p. 8). This analogy bears on the roles of *abstraction* and *lowered expectations* in science. Weinberg's (1976) description of the Galilean style emphasizes the *abstraction* of ordinary phenomena, in the process distancing oneself from commonsense ideas about them (as is routine in physics). Chomsky is thus analogizing the Cartesian distinction between body and soul (i.e., mind)—only the "body" can be understood—and a science of physics that *lowers its expectations* for understanding natural phenomena.[14] He suggests that physics may be "a remarkable historical accident resulting from chance convergence of biological properties of the human mind with some aspect of the real world" (Chomsky, 1980, p. 9).

As Section 4.2 argues, the intractability of CALU places limits on a science of language, in this way complementing the Galilean lowering of expectations for scientific understanding. In the biolinguistic framework, ordinary language use is segmented off from its underlying competence out of a recognition that biological properties of the human mind do *not* appear to attain a 'chance convergence' with this aspect of reality. This is why Chomsky deems the Galilean style to be a "narrower version" of the Cartesian thesis on free human action (Chomsky, 1980, p. 8)—it is a complementary base of inquiry from which a competence-performance distinction is derived, though never losing sight of the internalist orientation of linguistic science.[15]

14) Settling instead, as Chomsky (2009b) elsewhere argues, for intelligible *theories* of natural phenomena, rather than intelligible *phenomena*.

15) As Collins (2023, pp. 6–9) explains, a difference between generative linguistics and Galileo's form of idealization is, despite early invocations about the ideal 'speaker-hearer,' that generative linguistics targets an internal state whereas Galileo sought explanations for motion and interactions in space.

Free linguistic behavior can be made *partly* scientifically tractable by studying its underlying mechanisms, dependent on theoretical accommodations including competence-performance and a willingness to abstract away from linguistic performance to an idealized competence. As Collins explains, Galileo sought to 'decompose' natural phenomena to diminish the influence of interaction effects "in the search of formally precise invariances" (Collins, 2023, p. 5). Many factors contribute to language use; the Galilean disposition toward idealization "is a way of isolating an invariant phenomenon from general factors…" (Collins, 2023, p. 6). Adhering to this, Chomsky abstracts away from the brain's performance systems (Rey, 2020, p. 20), thereby conceiving of language *as it would be* in the absence of interaction effects. This reflects not an attitude towards data *per se*, but an attitude towards *abstraction* which informs data-selection (Collins, 2023, pp. 2–3). Piantadosi, in conceiving of LLMs-as-theories that refute the search for underlying principles at the expense of broad data coverage, misses this step and the analogy with CALU.

5.3 LLMs-As-Theories Re-Evaluated

LLMs are not language users because they do not replicate CALU. As a result, unless one imposes theoretical tenets onto the model, LLMs do not by themselves *identify*, *explain*, or shed useful *theoretical light* on CALU in humans. If LLMs are theories of human language, on what basis do they not deal with its creative uses? On what basis do they eschew a competence-performance distinction and the abstraction to language's enabling mechanisms? On what basis is language use *scientifically tractable* from the standpoint of broad data coverage? It is not clear that the statistical association of broad data offers *any* explanatory insight into CALU or aids the construction of a theory that makes it partly tractable.

This does not rule out the use of LLMs in linguistics, albeit in more limited respects. LLMs may demonstrate that there are *more regularities* in natural language use than our intuitions would lead us to expect. Yet, LLMs cannot lay claim to the fundamental facts about creative language use that can and should inform theory-construction—and already do within the biolinguistic approach.

One objection is that we are promoting a Galilean disposition toward idealization while also having accepted above, as Dupre (2024) argues, that computational modeling idealizes *too much*. The objection points us in a fruitful direction: language use pairs sound (or sign) with meaning and thus interfaces with other cognitive systems. Idealization, then, cannot ignore this non-linguistic cognitive activity. This raises questions about cognitive architecture, to which we now turn.

6 CALU, Computability, and Cognitive Architecture

As described, Mahowald et al. (2024) argue that LLMs achieve formal linguistic competence (knowledge of abstract rules and statistical regularities) but do not achieve functional linguistic competence (its use in real-world settings). This distinction is justified in reference to the *dissociation* of language and thought. The underlying problem is in their discussion of how LLMs or future artificial intelligence (AI) systems might achieve functional linguistic competence: through a *modular architecture* with parallels to the architecture of the human mind/brain. Mahowald et al. (2024, pp. 533–534) propose either *architectural* modularity or *emergent* modularity as the path to achieving both forms of competence—the explicit building of modularity into the architecture of a system or a natural induction of modularity through the training process,[16] respectively. Yet, no discussion of CALU *in humans* is provided in attaining this functional competence.[17]

The question is one of cognitive architecture. The biolinguistic framework already lays claim to this problem: its "chief hypothesis" is that language is "subserved by a language faculty, a computational system that, abstractly specified, realizes a function or procedure that generates structures (syntax) that encode the properties that allow a speaker hearer to pair sign…with meaning" (Collins, 2023, p. 1). Attempts to leverage LLMs as Mahowald et al. (2024) do only heighten the need for a proper conceptualization of CALU's cognitive basis beyond the internalist orientation and competence-performance distinction illustrated above. Let us take each component of CALU to see how the postulation of a generative language faculty accounts for the reformulated Cartesian problem and then contrast this with the use of LLMs as models of human language.

6.1 Unboundedness and the Language Faculty

The biolinguistic conception of a generative language faculty has roots in the intertwined history of modern Universal Grammar (UG) and computability theory. The language faculty is *assumed* to exist under UG, reflecting the history in which UG was "revived without awareness in the generative enterprise" with the new conceptual tools provided by computability theory (Chomsky, 2021, p. 9). Language, under UG, is a natural object characterized by *recursive enumerability* which yields *digital infinity* (Mendívil-Giró, 2018, p. 861). The postulation of a computational system that recursively produces grammatical sentences over an indefinite range is made on the basis that a brain with finite memory cannot list all possible grammatical sentences (Mendívil-Giró, 2018, pp. 873–874). The term *recursion* in modern UG's early days was directly inspired

16) This includes the training data and the objective function.

17) Mahowald et al. (2024, pp. 535-536) carefully note that whether LLMs are language users is one of several outstanding questions. Yet, CALU is unduly neglected.

PsychOpen GOLD

by mathematical logic, making "*recursive*...equivalent to *computable*" (Mendívil-Giró, 2018, p. 874). Indeed, the Minimalist Program's "Merge" operation preserves recursion as a property of a computational system (Mendívil-Giró, 2018, pp. 874–875).

Huybregts (2019) echoes this reasoning and relates it to the derivation of the competence-performance distinction, where competence is conceived as a computational system:

> Recursiveness is a property of the generative procedure applicable to any input, not a feature of its output, which may be arbitrarily constrained by complicating idiosyncratic factors independent of the procedure. The procedure may generate an infinite language but only produce a finite subset of it (Huybregts, 2019, p. 3).

Thus, a distinction between linguistic competence and performance is the *null hypothesis*, not vice versa (Huybregts, 2019, p. 4). Huybregts' broader argument—that infinite generativity cannot be reached via a stepwise approach—is relevant to the postulation of a competence characterized by a computational system, though one made only *after* the Cartesian problem is reformulated in the biolinguistic framework. The generative language faculty enables the unboundedness of human language use.

6.2 Stimulus-Freedom and the Language Faculty

Stimulus-freedom raises its own problem: How does language 'regulate' the generation of form-meaning pairs in the absence of identifiable stimuli? This problem goes beyond simplistic input-output processing: it is the problem of 'free access' to language in which linguistic resources can be recruited effectively *at will* (Collins, 2004, pp. 519–520). Crucially, the biolinguistic framework does not attempt to solve *this* problem, but to identify the cognitive mechanism(s) that—at least in part—*enable* this free access.

Based on our foregoing analysis, the answer is *internal*—the enabling mechanism will not be found in the individual's environment, but rather 'inside the head.' Language's cognitive basis must possess a degree of operational autonomy (McGilvray, 2005, p. 222). Specifically, whatever enables stimulus-freedom must fulfill two responsibilities: (1) the production of structured expressions with a discrete set of cognitive resources; and (2) interfacing with other, relevant cognitive systems. McGilvray explains that a *modular language faculty* is required:

> The possibility of stimulus freedom in language use can be seen to result from a modular language faculty. Not just any kind of modularity will do: we need a faculty that utilizes its own resources and with internal prompting produces (in apparent isolation) through its own algorithms items in the form of linguistic expressions that are unique to it and yet that "interface" with relevant other internal biological systems (McGilvray, 2001, p. 12).

The emphasis is on the *modularity* of a *generative language faculty*; a system in possession of "conceptual resources" that can make these resources "legible" to others in the mind (McGilvray, 2005, p. 218). The *modularity* of the language faculty merely indicates that this system can be studied independently of others for the sake of explanatory theory, in this case with its own domain-specific resources (see McGilvray, 2014, p. 235), though "as part of a broader investigation of its interactions with other such systems..." (Chomsky, 2013, p. 35).

One could eschew the role of a domain-specific language faculty in favor of, say, the interactivity of domain-general mechanisms. However, if CALU depended merely on domain-general mechanisms, then one *should* expect to find this phenomenon in the purposeful behaviors of non-human animals. Yet, we do not, making this conclusion dubious (Baker, 2007, pp. 239–240).

6.3 Appropriateness, Abduction, and Modularity

To get at its cognitive basis, Baker (2007) poses CALU as a *poverty of stimulus* problem. To acquire CALU, a child would have to learn that its parents are *not* automata because they use their language in a stimulus-free, unbounded, and appropriate manner. The child would then have to learn that *they too* can use language this way. The third step is where the effort at empirical analysis appears to collapse: the child would then have to learn "*how* not to be an automaton—how to develop the capacity to use language in this way" (Baker, 2007, p. 241). This cannot even be *framed* properly, as "we have no precise algorithmic way to specify the knowledge that this capacity depends on or the processes that it involves" and therefore "we cannot estimate the amount of information that is involved" (Baker, 2007, p. 241). Baker (2007, pp. 241–245) goes on to argue that the more basic problem of determining whether those around the child are automata is not learnable, either. CALU, then, is plausibly innate.

Baker's (2007, pp. 239–240) (tentative) conclusion about its cognitive basis is notable: that CALU is a module of the human mind[18]—though, not localized in a particular brain region (Baker, 2011, pp. 90–91). Baker comes to a parallel conclusion: that the CALU module cannot be characterized computationally. This conclusion turns on the observation that "the whole notion of "appropriate" is an abductive one. We judge that what someone says to us is appropriate not at all on the basis of the syntactic structure of what is said, but entirely on the semantic properties of what is said" (Baker, 2007, p. 253). Baker's (2007, pp. 252–253) implication, drawing from Fodor (2000), is that it makes sense CALU has not (as of writing) been replicated via a computer program, as abductive reasoning seems beyond the scope of computation.

18) Baker (2007, p. 240) carefully notes that this is "at least for the sake of argument."

6.4 LLMs, CALU, and Architecture

Using LLMs as models of human language with parallels to cognitive architecture runs into two problems. First, attempts to leverage LLMs as models of human language risk "[reversing] the nature of the argument" (Leivada, Dentella, & Günther, 2024, p. 3). Any scientific account of human language must deal with CALU. Yet, arguing that LLMs are useful models (or theories) of human linguistic competence without incorporating the Cartesian observations is to start from an incomplete point of inquiry. Biolinguistics recognizes the Cartesian problem, splits it for the sake of scientific tractability, and uses the tools that enable this split—in part, computability theory—to postulate an innate computational system that makes possible creative language use. To conceive of LLMs as mastering the abstract structure of natural language is entirely premature as they offer no conception of the problem whatsoever. Even for Mahowald et al. (2024) to ground their formal-functional distinction in neuroscientific data about their respective mechanisms ignores the insight that "[d]ata only have meaning within a theoretical framework…If the framework is deficient, the interpretation of the data will be insufficient, too" (Mahlmann, 2023, p. 346).

Second, the attempt to attain functional competence in LLMs (or future AI systems) with comparison to human language use is similarly deficient in its foundations. As Section 5.2 illustrates, there is a deliberate *lowering of expectations* in the biolinguistic approach out of a recognition of CALU's scientific intractability. Discussions of LLMs' functional competence understate the *seriousness* of the problem of language use, hardly recognizing—if at all—its creative character. As a result, science is given full steam to study problems it may be unable to penetrate, or seriously unlikely barring theoretical accommodations that LLMs may not permit or indicate (e.g., competence-performance, idealization of the language faculty, etc.).

Consider how Transformer-based LLMs do not replicate CALU—they do not replicate its stimulus-freedom, appropriateness, and the strong interpretation of unboundedness. This plausibly rules out the Transformer architecture in the pursuit of functional competence, as achieving these would require changes of a fundamental nature. In principle, however, the mechanisms that enable, at least, stimulus-freedom and unboundedness are *computable* and can thus be reproduced via a different architecture. This would likely require *architectural* modularity—explicit building of modules.

Appropriateness is a trickier matter. In Section 2.2, we identified that appropriateness is not *merely* functional and the stimulus-controlled nature of LLM-powered chatbots, fine-tuned via RLHF, is *purely* functional. This could plausibly be resolved, though not by a system with a "conception" of appropriateness provided via RLHF, as this is premised on a functional system/end-user dynamic. How a system *could be* designed to achieve this is a more difficult question. RLHF is a data-driven process, relying on structured and explicit examples of human-generated data of what counts as (in)appropriate outputs—and this is not even sufficient. Yet, the only way to scientifically *make sense of* CALU is internalism; the cognitive mechanisms that enable it. We have no theory that provides a list of all the circumstances and associated utterances that are appropriate, as CALU indicates such a theory may be a fool's errand (see McGilvray, 2001, pp. 8–18).

We are forced back into the mind. Baker (2007) argued that interpreting and producing appropriate expressions appears to be an *abductive* process and modular at that. Research on representing abduction computationally has progressed. However, computational tractability—an efficient algorithmic solution—remains a roadblock (see van Rooij et al., 2019, ch. 12). Human brains possess limited computational power, indicating that brute force is not the key to tractability (see Blokpoel et al., 2018, p. 2). A solution, then, is likely to require explicit building of modules into a system whose architecture is radically different than that of a Transformer-based system.

All this assumes that computation *alone* supports CALU. This assumption has its doubters (e.g., Baker, 2007, pp. 252–253). It is not clear how computation can direct itself, so to speak, in a manner that escapes the bounds of the input it receives. Here, we get to the heart of the biolinguistic approach to CALU, the value of splitting the problem for the sake of scientific tractability, and the usefulness of lowering scientific expectations. We have been exploring each component of CALU *individually*; stimulus-freedom, unboundedness, and appropriateness. But CALU does not exist in this carved-up fashion—language use simply *is* free. When we break it down in this way we are doing so as part of an exercise to theoretically accommodate it. It is not clear if—and how—the mechanisms for each component being computable would collectively give rise to CALU in a computational model (in this way making it, presumably, something more than just a model). The split of the Cartesian problem is a *theoretical accommodation*; the cognitive mechanisms that make CALU possible are a necessary *but perhaps not sufficient* condition for CALU to arise. CALU arises under some configuration of these three components, but there is no logical certainty that they are sufficient (see Chomsky, 1982, p. 433). This raises doubts about the possibility of replicating CALU on a computational device, architectural or emergent modularity notwithstanding.

All this illustrates the extent to which Mahowald et al. (2024) overstate the success of LLMs and what they forebode. The dissociation of language and thought has significant negative effects on explanatory theory-construction, chiefly in the deficient characterization of ordinary human language use, the lack of construction of a theoretical framework to make CALU partly tractable, and a failure to lay claim to a cognitive architecture that enables—but does not explain—free language use. Whereas some criticize biolinguistics from the standpoint of computational modeling for exhibiting a 'contempt for applications' of theoretical work (Pullum, 2009, p. 17), the approach recognizes the difference between *explanation* and *application*; it recognizes the scale and distinctiveness of the problem of CALU and the need to make *part of it* scientifically tractable.

7 Conclusion

The Cartesian problem of creative language use presents a set of facts that must be dealt with by a science of language. It is dealt with in the biolinguistic framework by identifying those mechanisms that, in part, make CALU possible, thereby splitting up the Cartesian problem for the sake of scientific tractability. To be sure, the postulation of a generative language faculty is "not logically necessary" (Anderson, 2008, p. 804). Conceiving of LLMs as theoretically significant, theories, or models of human language *is* logically possible. Yet, linguistic theory premised on LLMs should be evaluated according to their "explanatory/unifying power" and ability to "do things previous approaches could not, but without discarding the reasons we found the prior approach plausible in the first place" (Dupre, 2024, p. 11). LLMs do not achieve this vis-à-vis CALU.

"It is not a novel insight," Chomsky writes, "that human speech is distinguished by these qualities, though it is an insight that must be recaptured time and time again" (Chomsky, 2006, p. 88). Critics of biolinguistics sometimes miss the implicit point: that the mind is not as amenable to scientific inquiry as we would like. The use of LLMs to explain human language has a way of underestimating the scale and quality of the problems it poses. This paper has made the case that CALU, one such problem, cannot be bypassed and LLMs do not offer a better path than the biolinguistic framework.

Funding: The author has no funding to report.

Acknowledgments: I am very grateful to Kleanthes Grohmann and the rest of the editorial team at *Biolinguistics* for their feedback and direction in handling this paper. I also wish to thank the three anonymous reviewers for insightful and constructive comments that dramatically improved the content of this paper. Finally, for their intellectual assistance and words of encouragement, I wish to thank Elliot Murphy, Mark Baker, and Stephen Pulman. Any mistakes are of course my own.

Competing Interests: The author has declared that no competing interests exist.

References

Abramson, D. (2011). Descartes' influence on Turing. *Studies in History and Philosophy of Science, 42*(4), 544–551. https://doi.org/10.1016/j.shpsa.2011.09.004

Anderson, S. (2008). The logical structure of linguistic theory. *Language, 84*(4), 795–814. https://doi.org/10.1353/lan.0.0075

Asoulin, E. (2013). The creative aspect of language use and the implications for linguistic science. *Biolinguistics, 7*, 228–248. https://doi.org/10.5964/bioling.8963

Baker, M. C. (2007). The creative aspect of language use and nonbiological nativism. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind* (Vol. 3, pp. 233–253). Oxford University Press.

Baker, M. C. (2011). Brains and souls; grammar and speaking. In M. C. Baker & S. Goetz (Eds.), *The soul hypothesis* (pp. 73–98). The Continuum International Publishing Group Inc.

Baroni, M. (2022). On the proper role of linguistically oriented deep net analysis in linguistic theorizing. In S. Lappin & J. Bernardy (Eds.), *Algebraic structures in natural language* (pp. 1–16). CRC Press.

Beckers, G. J. L., Huybregts, M. A. C., Everaert, M. B. H., & Bolhuis, J. J. (2024). No evidence for language syntax in songbird vocalizations. *Frontiers in Psychology, 15*, Article 1393895. https://doi.org/10.3389/fpsyg.2024.1393895

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Berwick, R. C., Pietroski, P., Yankama, B., & Chomsky, N. (2011). Poverty of the stimulus revisited. *Cognitive Science, 35*(7), 1207–1242. https://doi.org/10.1111/j.1551-6709.2011.01189.x

Blokpoel, M., Wareham, T., Haselager, P., Toni, I., & van Rooij, I. (2018). Deep Analogical inference as the origin of hypotheses. *The Journal of Problem Solving, 11*(1), Article 3. https://doi.org/10.7771/1932-6246.1197

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... (2020). Language models are few-shot learners. *ArXiv*. https://arxiv.org/abs/2005.14165v4

Chomsky, N. (1959). A review of B. F. Skinner's Verbal Behavior. *Language, 35*(1), 26–58. https://doi.org/10.2307/411334

Chomsky, N. (1964). *Current issues in linguistic theory*. Mouton and Co.

Chomsky, N. (1966). *Cartesian linguistics*. Harper & Row.

Chomsky, N. (1975). Knowledge of language. In K. Gunderson (Ed.), *Language, mind, and knowledge* (pp. 299–320). University of Minnesota Press.

Chomsky, N. (1980). *Rules and representations*. Columbia University Press.

Chomsky, N. (1982). A note on the creative aspect of language use. *The Philosophical Review, 91*(3), 423–434. https://doi.org/10.2307/2184692

Chomsky, N. (1988). *Language and problems of knowledge*. The MIT Press.

Chomsky, N. (1995). Language and Nature. *Mind, 104*(413), 1–61. https://doi.org/10.1093/mind/104.413.1

Chomsky, N. (2006). *Language and mind* (3rd ed.). Cambridge University Press.

Chomsky, N. (2009a). *Cartesian linguistics*. Cambridge University Press.

Chomsky, N. (2009b). The mysteries of nature. *The Journal of Philosophy, 106*(4), 167–200. https://doi.org/10.5840/jphil2009106416

Chomsky, N. (2009c). Turing on the "Imitation Game." In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the Turing Test* (pp. 103–106). Springer Netherlands.

Chomsky, N. (2013). Problems of projection. *Lingua, 130*, 33–49. https://doi.org/10.1016/j.lingua.2012.12.003

Chomsky, N. (2017). The Galilean Challenge. *Inference-Review, 3*(1), 1–7. https://inference-review.com/article/the-galilean-challenge

Chomsky, N. (2021). Linguistics then and now. *Annual Review of Linguistics, 7*(1), 1–11. https://doi.org/10.1146/annurev-linguistics-081720-111352

Chomsky, N., & Moro, A. (2022). *The secrets of words.* The MIT Press.

Collins, J. (2004). Faculty disputes. *Mind & Language, 19*(5), 503–533. https://doi.org/10.1111/j.0268-1064.2004.00270.x

Collins, J. (2021). Chomsky's problem/mystery distinction. In N. Allot, T. Lohndal, & G. Rey (Eds.), *A companion to Chomsky* (pp. 557–566). John Wiley & Sons.

Collins, J. (2023). Generative linguistics. *Language Sciences, 100*, Article 101585. https://doi.org/10.1016/j.langsci.2023.101585

Cohen, I. B. (1985). *Revolution in science.* Harvard University Press.

Contreras Kallens, P., Kristensen-McLachlan, R. D., & Christiansen, M. H. (2023). Large Language models demonstrate the potential of statistical learning in language. *Cognitive Science, 47*(3), Article e13256. https://doi.org/10.1111/cogs.13256

Cordemoy, G. d. (1668). *A philosophicall discourse concerning speech, conformable to the cartesian principles.* The British Library.

den Ouden, B. D. (1975). *Language and creativity.* The Peter de Ridder Press.

Dentella, V., Günther, F., & Leivada, E. (2023). Systematic testing of three Language Models reveals low language accuracy, absence of response stability, and a yes-response bias. *Proceedings of the National Academy of Sciences of the United States of America, 120*(51), Article e2309583120. https://doi.org/10.1073/pnas.2309583120

Descartes, R. (1910/1637). *Discourse on the method of rightly conducting the reason, and seeking truth in the sciences.* The Open Court Publishing Company.

Dupre, G. (2024). Acquiring a language vs. inducing a grammar. *Cognition, 247*Article 105771. https://doi.org/10.1016/j.cognition.2024.105771

Fodor, J. (2000). *The mind doesn't work that way.* The MIT Press.

Fox, D., & Katzir, R. (2024). Large Language Models and theoretical linguistics. *Theoretical Linguistics, 50*(1-2), 71–76. https://doi.org/10.1515/tl-2024-2005

Galileo, G. (2001/1632). *Dialogue concerning the two chief world systems.* (S. Drake, Trans.). Modern Library.

Greco, M., Cometa, A., Artoni, F., Frank, R., & Moro, A. (2023). False perspectives on human language. *Frontiers in Language Sciences, 2*, Article 1178932. https://doi.org/10.3389/flang.2023.1178932

Gunderson, K. (1964). Descartes, La Mettrie, language, and machines. *Philosophy, 39*(149), 193–222. https://doi.org/10.1017/S0031819100055595

Hu, J., Mahowald, K., Lupyan, G., & Levy, R. (2024). Language models align with human judgments on key grammatical constructions. *PNAS, 121*(36), Article e2400917121. https://doi.org/10.1073/pnas.2400917121

Huybregts, M. A. C. (2019). Infinite generation of language unreachable from a stepwise approach. *Frontiers in Psychology, 10*(425), Article 425. https://doi.org/10.3389/fpsyg.2019.00425

Jones, C. R., & Bergen, B. K. (2024). People cannot distinguish GPT-4 from a human in a Turing Test. *ArXiv*. https://doi.org/10.48550/arXiv.2405.08007

Kallini, J., Papadimitriou, I., Futrell, R., Mahowald, K., & Potts, C. (2024). Mission: Impossible Language Models. *ArXiv*. https://doi.org/10.18653/v1/2024.acl-long.787

Katzir, R. (2023). Why Large Language Models are poor theories of human linguistic cognition. *Biolinguistics, 17*, Article e13153. https://doi.org/10.5964/bioling.13153

Kodner, J., Payne, S., & Heinz, J. (2023). *Why linguistics will thrive in the 21st century.* https://lingbuzz.net/lingbuzz/007485

Lakretz, Y., Hupkes, D., Vergallito, A., Marelli, M., Baroni, M., & Dehaene, S. (2021). Mechanisms for handling nested dependencies in Neural-Network language models and humans. *Cognition, 213*(3), Article 104699. https://doi.org/10.1016/j.cognition.2021.104699

Land, S. K. (1974). The Cartesian Language Test and Professor Chomsky. *Linguistics, 12*(122), 11–24. https://doi.org/10.1515/ling.1974.12.122.11

Leivada, E., Dentella, V., & Günther, F. (2024). Evaluating the Language abilities of Large Language Models vs. humans. *Biolinguistics, 18*, Article e14391. https://doi.org/10.5964/bioling.14391

Leivada, E., Dentella, V., & Murphy, E. (2024). The quo vadis of the relationship between language and Large Language Models. *ArXiv*. https://doi.org/10.48550/arXiv.2310.11146

Mahlmann, M. (2023). *Mind and rights.* Cambridge University Press.

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2024). Dissociating language and thought in large language models. *Trends in Cognitive Sciences, 28*(6), 517–540. https://doi.org/10.1016/j.tics.2024.01.011

McCoy, R. T., Smolensky, P., Linzen, T., Gao, J., & Celikyilmaz, A. (2023). How much do language models copy from their training data? *Transactions of the Association for Computational Linguistics, 11*, 652–670. https://doi.org/10.1162/tacl_a_00567

McGilvray, J. (1999). *Chomsky.* Polity Press.

McGilvray, J. (2001). Chomsky on the creative aspect of language use and its implications for lexical semantic studies. In F. Busa & P. Bouillon (Eds.), *The language of word and meaning* (pp. 5–27). Cambridge University Press.

McGilvray, J. (2005). Meaning and creativity. In J. McGilvray (Ed.), *The Cambridge companion to Chomsky* (pp. 204–222). Cambridge University Press.

McGilvray, J. (2009). *Introduction to Cartesian linguistics.* Cambridge University Press.

McGilvray, J. (2014). *Chomsky.* Polity Press.

McGinn, C. (1993). *Problems in philosophy.* Blackwell Publishers.

Mendívil-Giró, J. L. (2018). Is Universal Grammar ready for retirement? *Journal of Linguistics, 54*(4), 859–888. https://doi.org/10.1017/S0022226718000166

Milway, D. (2023). *A response to Piantadosi.* https://lingbuzz.net/lingbuzz/007264

Moro, A. (2016). *I speak, therefore I am.* Columbia University Press.

Moro, A., Greco, M., & Cappa, S. F. (2023). Large languages, impossible languages, and human brains. *Cortex, 167*, 82–85. https://doi.org/10.1016/j.cortex.2023.07.003

Murphy, E., & Leivada, E. (2022). A model for learning strings is not a model of language. *Proceedings of the National Academy of Sciences of the United States of America, 119*(23), Article e2201651119. https://doi.org/10.1073/pnas.2201651119

OpenAI. (2022, November 30). Introducing ChatGPT. *OpenAI.* Accessed January 26, 2024, https://openai.com/blog/chatgpt

OpenAI. (2023). *GPT-4 technical report.* ArXiv. https://doi.org/10.48550/arXiv.2303.08774

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). *Training language models to follow instructions with human feedback.* ArXiv. https://arxiv.org/abs/2203.02155v1

Piantadosi, S. (2023). *Modern language models refute Chomsky's approach to language.* https://lingbuzz.net/lingbuzz/007180

Pullum, G. K. (2009). Computational linguistics and generative linguistics. *Proceedings of the EACL 2009 Workshop on the Interaction Between Linguistics and Computational Linguistics*, 12–21. https://dl.acm.org/doi/10.5555/1642038.1642042

Pulman, S. (2018, June 13). Language, learning and creativity. *University of Cambridge, Annual Wheeler Lectures.* Accessed June 29, 2024, from https://sms.cam.ac.uk/media/2770009.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training*. OpenAI. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners.* OpenAI. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Rees, T. (2022). Non-human words. *Daedalus, 151*(2), 168–182. https://doi.org/10.1162/daed_a_01908

Riskin, J. (2016). *The restless clock.* University of Chicago Press.

Rey, G. (2020). *Representation of language.* Oxford University Press.

Rosenfield, L. C. (1968). *From beast-machine to man-machine* (New and enlarged ed.). Octagon Books.

Shanahan, M. (2023). *Talking about Large Language Models.* ArXiv. https://arxiv.org/abs/2212.03551v5

Skinner, B. F. (1957). *Verbal behavior.* Copley Publishing Group.

Smith, N. (2004). *Chomsky.* Cambridge University Press.

Spelke, E. (2010). Innateness, choice, and language. In J. Franck & J. Bricmont (Eds.), *Chomsky notebook* (pp. 203–210). Columbia University Press.

Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H.-T., Jin, A., Bos, T., Baker, L., Du, Y., Li, Y., Lee, H., Zheng, H. S., Ghafouri, A., Menegali, M., Huang, Y., Krikun, M., Lepikhin, D., Qin, J., … (2022). *LaMDA.* ArXiv. https://arxiv.org/abs/2201.08239v1

Tomalin, M. (2006). *Linguistics and the formal sciences.* Cambridge University Press.

Turing, A. M. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, s2-42*(1), 230–265. https://doi.org/10.1112/plms/s2-42.1.230

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

van Rooij, I., Blokpoel, M., Kwisthout, J., & Wareham, T. (2019). *Cognition and intractability.* Cambridge University Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention is all you need.* ArXiv. https://arxiv.org/abs/1706.03762v1.

Warstadt, A., Singh, A., & Bowman, S. R. (2019). Neural network acceptability judgments. *Transactions of the Association for Computational Linguistics, 7*(1), 625–641. https://doi.org/10.1162/tacl_a_00290

Watumull, J., & Chomsky, N. (2020). Rethinking universality. In A. Bárány, T. Biberauer, J. Douglas, & S. Vikner (Eds.), *Syntactic architecture and its consequences II* (pp. 3–24). Language Science Press.

Weinberg, S. (1976). The forces of nature. *Bulletin of the American Academy of Arts and Sciences, 29*(4), 13–29. https://doi.org/10.2307/3823787

Wilcox, E. G., Futrell, R., & Levy, R. (2024). Using computational models to test syntactic learnability. *Linguistic Inquiry, 55*(4), 805–848. https://doi.org/10.1162/ling_a_00491

Ye, H., Liu, T., Hua, W., & Jia, W. (2023). *Cognitive mirage.* ArXiv. https://doi.org/10.48550/arXiv.2309.06794