On Modern Language Models, Impossible Languages, and Anti-science

Dan Milway

August 2, 2024

The proponent of a new theory has a distinct advantage against the defender of the well-established theory that the former seeks to upend. Namely, the one who would propose a something new can know every argument in the other's repertoire, while the reverse cannot be true. For instance in his *Dialogue*, Galileo points out that every Copernican knew the arguments of Aristotle, Ptolemy, and their followers, but none of the Aristotelians or Ptolemaics knew the arguments of the Copernicans (Galilei, 1632/2001, pp. 148–149).

With the recent fervour around Modern Language Models (MLMs)—largely developed by for-profit corporations—anti-generativists have declared us to be in such a moment. Piandadosi (2023), for instance writes that "[a]fter decades of privilege and prominence in linguistics, Noam Chomsky's approach to the science of language is experiencing a remarkable downfall." (p1) For Piantadosi, and those whose work he highlights, Chomsky's theories are the status quo and they, as the outsiders, are in prime position to take it out. The arguments, evidence, and internal disputes of the Generative enterprise are widely published and taught within the academy and without[1].

Despite this, though, Piandadosi (2023) demonstrates almost no knowledge of either the theoretical or empirical content of Generative linguistics. Indeed, even when directly confronted with empirical arguments in favour of generative grammar, Piantadosi simply ignores them. For instance, when attempting to refute the claim that MLMs can learn impossible languages—*i.e.*, those that no human can acquire —the proponents of MLMs fail to engage with the empirical basis for impossible languages. As a result, the attempted refutations of the impossible languages critique are toothless at best and support the critique at worst. I discuss this in detail in section 1.

Worse still, Piantadosi's polemic demonstrates a surprising amount of disdain for the aims, history, and norms of science. In it, he disparages explanatory theories, not just in linguistics but in the sciences more broadly, in favour of black box prediction machines like the MLMs he promotes. This, I argue in section 2, stems from a misunderstanding of scientific explanation and a very low bar for predictiveness.

1 On Impossible Languages

A recurring generativist critique of MLMs *qua* linguistic theories is the proposition that they are capable of learning impossible languages—*i.e.*, languages which humans are incapable of acquiring (Chomsky et al., 2023; Moro et al., 2023). One could imagine two potential but contradictory way to refute this

[1]Michel Gondry's 2013 film *Is the man who is tall happy?* features a non-technical discussion of Chomsky's linguistics

critique—either dispute the very idea that there are impossible languages of the sort that generativists assert, or show that MLMs, like humans, cannot learn impossible languages . The proponents of MLMs that do address impossible languages (Kallini et al., 2024; Piandadosi, 2023), as we shall see, oddly seem to make both arguments. I will further argue that both responses misconstrue the very idea of impossible languages, leading to flawed arguments and that this misconstrual stems from a failure to recognize foundational differences between generativist theories of grammar and MLMs. First, though, we must lay out what generativists mean when we talk about impossible languages.

1.1 What Are Impossible Languages?

There are a number of ways that a language could be considered impossible or unlearnable that are obvious without empirical or theoretical linguistic investigations. A language with infinite length sentences is practically impossible. A languages that includes a random shuffle as a grammatical transformation/rule is unlearnable from an information-theoretic stand-point. Decades of study by generative linguists, though, have revealed classes of of languages which, though they are perfectly plausible in the abstract, are not humanly possible—could not be acquired as actual languages are acquired. It is this latter class of languages that is at issue in the present debate.

Contemporary generative linguistic theory consists roughly of the following propositions.

1. The human language faculty, narrowly construed (FLN) is a recursive procedure that generates an infinite array of structured expressions.
2. The sole structure building operation of FLN is Merge.
3. Merge is binary set formation (Merge(α, β) $\rightarrow \{\alpha, \beta\}$)
4. Merge as two possible cases: Internal (α is a term of β or vice versa), and External (neither α nor β is a term of the other).

It is important to note that, because Merge builds structures without linear order, generative theory predicts that grammatical rules and intuitions may be sensitive to structural configurations, but not linear order. Thus, a sentence like (1) is interpreted such that the adverb *instinctively* modifies the closest verb in terms of structural distance—*swim*—rather than linear distance—*fly*.

(1) Instinctively, eagles that fly swim. (Chomsky, 2013)

Furthermore, this is a universal principle—there are no possible human languages with syntactic rules based on linear order rather than hierarchical structure.

We could, of course construct a language, call it L-English, in which modification is sensitive to linear rather than structural distance. In such a language, for instance, (1) would mean that eagles that instinctively fly, also swim. According to generative theory, though, L-English is an impossible language and no human child could acquire it as they would English regardless of the amount and type of data received. The respective grammars of L-English and English are different *kinds* of entity.

Being a science, generative linguistics need not stop at speculation—it can test these predictions, and indeed it has. Smith and Tsimpli (1995) present a case-study of a "savant" linguist who can seemingly learn any possible language with ease but cannot learn an impossible language. While these results are revealing, they are limited by the fact that they are gleaned from the study of one individual. More recently, though, Musso et al. (2003) and Tettamanti et al. (2002) have tested the impossible language prediction under more standard experimental condiditons, providing neural and behavioural data from multiple neurotypical adults. Both studies reach the conclusion that the possible and impossible languages respectively are learned by different parts of the brain.[2] That is, they support the prediction.

This body of research yields a test for theories or models which purport to explain, simulate, or duplicate the human language faculty. Put in terms of MLMs, if a language model can learn both possible and impossible languages in the same way—if it treats possible and impossible languages as the same sort of things, then it cannot be an accurate model of the human language faculty. Unfortuantely, the two papers which seem to address the notion of impossible languages, fail to understand it, and thus are unconvincing.

1.2 Can MLMs learn impossible languages?

Piantadosi, who is a non-generativist, can perhaps be forgiven for failing to address the impossible languages in his opening salvo,[3] but his response to this evidence being pointed out to him (Milway, 2023; Moro et al., 2023) is truly damning.

His first response is to tacitly admit that MLMs can learn impossible languages, musing that "nobody seems to think its important to first know how much data these architectures (or nearby ones) would require in order to learn 'possible' versus 'impossible' languages." (36) The implication here is that if MLMs can learn impossible languages but can only do so with an unreasonably large set of training data that that is just as good as not being able to learn impossible languages. Two objections to such an assertion leap immediately to mind.

The first is one that is orthogonal to the impossible language critique. It is that inversely correlating the amount of a model's training data with judgments of that model's performance—*i.e.*, the greater amount of training data required, the worse the performance of a model—leads to precisely the opposite conclusions of Piantadosi's. It would mean that since, by Kodner et al.'s (2023) estimate, GPT4 requires training data equivalent to the PLD of millions of human learners, it performs much worse than a human infant. What's more, since the trend seems to be moving toward larger amounts of training data (Kaplan et al., 2020), MLMs are getting worse.

The second objection is that it is question-begging—it only follows if one incorrectly treats the possible–impossible dichotomy as a difference in degree rather than one of kind. It is as if in response to an assertion that Model X is not a good model of number theory, because it generates spurious factorizations of prime numbers, a proponent of Model X says "but it does so with difficulty."

His second response is to question the very notion of impossible languages, writing that "the idea of 'impossible languages' has never actually been empirically justified. Nobody knows what the space of possible languages is. There are no examples where universals have been shown to be due to something specific to language through rigorous comparison to non-linguistic domains." (37) This is simply factually untrue, and had Piantadosi made the barest of good-faith efforts of reviewing the relevant literature —*i.e.*, the literature discussed in section 1.1 before responding, he would know that it's factually untrue.[4]

[2] "This stands as neurophysiological evidence that the acquisition of new linguistic competence in adults involves a brain system that is different from that involved in learning grammar rules that violate UG." (Musso et al., 2003, pp. 777–778)

"[T]his study demonstrates a selective and robust participation of Broca's area in the acquisition of novel [grammatical] rules as opposed to [nongrammatical] rules." (Tettamanti et al., 2002, p. 707)

[3] Given Piantadosi's intial claim of generative grammar's dominance and prestige the fact that he voluntarily chose to write its death-knell though, this is an exceedingly charitable position to take.

[4] Instead, he points to the much-ballyhooed example of Pirahã—fitting since the paper appears in a Festschrift for Daniel

Rather than attempting to dismiss the critique out of hand, Kallini et al. (2024) take it for the empirical challenge that it is, and attempt to test it. Responding directly to Chomsky et al.'s (2023) assertion that MLMs can acquire possible and impossible languages "with equal facility," Kallini et al. (2024) identify structural dependency as the predicted limit of possible languages. Kallini et al. (2024) first construct what they call an *Impossibility Continuum*—a sequence of perturbation operations designed render an actual language (English) successively less possible. There are three classes of perturbations investigated ordered from most impossible to least impossible:

1. *Shuffle languages involve an arbitrary reordering of words/tokens.

2. *Reverse languages involve putting words/tokens in reverse order.

3. *Hop languages involve shifting subject-verb agreement affixes 4 words/tokens to the right.

Each class also includes a control language (NoShuffle, NoReverse, NoHop) which is essentially English. They then trained a variant of the GPT-2[5] transformer on the various perturbed datasets and measured the transformer's learning performance.

According to Kallini et al.'s (2024) results, the GPT-2 transformer has greater difficulty learning as it goes up the impossibility continuum. This is taken as a refutation of the impossible languages critique but it is based on a misrepresentation of the the critique—it takes (im)possibility to be a difference in degree rather than kind. The results, then, are inconclusive. Curiously, though, Kallini et al. (2024) present two relevant results in their appendices, and these results support the impossible languages critique.

The first experiment compares how well models trained on *Reverse and *Hop languages can identify constituents. They found that all of the models in question were indistinguishable from each other regardless of where the languages lay on the impossibility continuum. This is exactly the sort of result that the impossible languages critique would predict—GPT-2 is treats all languages alike regardless of what kind of language they are.

The second experiment looks at the models trained on the DeterministicShuffle languages, whose perturbation operation is described as follows.

> The tokenized input sentence is deterministically shuffled based on the length of the token sequence. For example, all token sequences of length 5 are shuffled in the same order. We create several languages by varying the random seed *s* that produces the deterministic shuffle. (Kallini et al., 2024, p. 4)

Kallini et al. (2024) used three DeterministicShuffle languages ($s \in \{21, 57, 84\}$). Because these shuffle algorithms are deterministic, they are in principle discoverable inductively from their output. The experiment under discussion tests the extent to which each model learned it's respective shuffle algorithm.

The results of this experiment are that each trained model was able to recognize the language it was trained on better than the other DeterministicShuffle languages and the NonDeterministicShuffle language. This suggests that, despite the DeterministicShuffle languages being high on Kallini et al.'s (2024}) impossibility continuum—*i.e.*, being a language that generativists and non-generativists might agree is impossible—GPT2 was able to learn these languages. Again, this supports the impossible language critique.

Everett—along with other examples of languages "which were argued not to have recursive embedding", without addressing or even acknowledging the fact that "All languages have recursive embedding" was never proposed as a universal.

[5]Developed by OpenAI inc.

So, while both Piandadosi (2023) and Kallini et al. (2024) claim to refute the critique that MLMs can learn impossible languages and are therefore ill-suited as models of human language, we have seen that their arguments are confused and open to complete dismissal and alternative interpretation respectively. Piandadosi (2023) seems to deliberately ignore the empirical basis for the critique, and so can be ignored. Kallini et al. (2024), though, seem to approach the critique in good faith, but their results are not a clear as they present them to be and in some cases, point to the opposite of their conclusions. Both responses, though, suffer from a failure to recognize the fundamental differences between themselves and the generativists to whom they are responding. I turn to these fundamental differences presently.

1.3 Foundational Differences

As I argue above, the possible–impossible language dichotomy predicted by generative theory is a difference of kind, yet both Piandadosi (2023) and Kallini et al. (2024) treat it as a difference in degree. This misinterpretation, though, does not seem to be deliberate. Rather, it stems from a fundamental and usually tacit conflict between rationalism and empiricism. Specifically, the very idea of an impossible language of the sort discussed by Chomsky et al. (2023) and Moro et al. (2023) entails a rationalist theory of language and mind, while MLMs are the products of a decades long empiricist project, making their proponents at the very least uneasy with it.

The distinction between the two philosophies lies fundamentally in how they each construe learning. Rationalism frames it as the learner applying innate mental ideas/structures to the data, while empiricism frames it as the data shaping the previously under-structured mind. It is not difficult to see how each view would lead to different notions of (im)possibility in the field of language. Both, of course, should be consistent with (im)possibility based on language-external factors—logic, physics, mathematics, *etc.*—but only rationalist theories could find languages which are impossible solely on the basis of the structure of the language faculty.

For rationalist linguists like Chomsky et al. (2023) and Moro et al. (2023), the innate structures of the human mind define our capacity to acquire language, with “capacity” referring to both our abilities and limitations. These structures are what I referred to above as FLN—often called UG. Humans can acquire a language only insofar as the language’s rules are consistent with UG. Furthermore, UG is but one component of our cognition and each other component will have its own innate structure that defines its capacity. This picture is borne out by the empirical results from Musso et al. (2003), Smith and Tsimpli (1995), and Tettamanti et al. (2002) which show that, to the extent that impossible languages are learnable, they are learnable, not by UG, but by some other component.

The empiricist position, on the other hand, is that there are no innate mental structures in the human mind. For the empiricist, the mind is purely a learning organ, and any structure that the mind has is acquired by the processing of sense data. There are two relevant consequences of this position. First, all knowledge is acquired inductively. A corollary of this is that there are no inherent connections between pieces of knowledge, only contingent associations. Second, the human mind can be reproduced once we discover the correct learning algorithm.

I have placed generativists like Chomsky and Moro into the rationalist camp because they tend to be quite explicit about their philosophical forebears—Moro et al. (2023), for instance, include a discussion of Descartes’ philosophy. Proponents of MLMs, on the other hand, seem to be less explicit about their philosophical influences—Piandadosi (2023) favourably cites philosophers Ned Block and Nelson Goodman in passing—yet the character of MLMs with their dependency on the size of their models and training datasets (Kaplan et al., 2020) is decidedly empiricist.

The empiricist bent of MLM's proponents is further supported by how they interpret the impossible language critique. As I argue above, one of the corollaries of empiricism is that no two ideas/concepts are inherently related. So, sharp distinctions like the one between possible and impossible languages are replaced with fuzzy ones. This can be seen by the fact that Piandadosi (2023) opines that the impossible languages merely require more data to learn or that Kallini et al. (2024) propose an impossibility continuum. Indeed, one could argue that the source of the problems with Piandadosi's and Kallini et al.'s arguments discussed in section 1.2 is that they interpret a rationalist argument in empiricist terms.

2 On Explanation and Prediction

A key difference between proponents of generative theories and those of MLMs is centers around the primacy of explanation vs prediction as goals of science. Generativists argue that science is first and foremost explanatory and that since MLMs are to complex to be properly understood they cannot be scientific theories. Piandadosi (2023) disputes this, arguing that prediction is more important and that, since MLMs have far greater predictive power than generative theories, the former supplants the latter. As we shall see, though, both premises of Piantadosi's arguments are flawed. In section 2.1, I will argue that the predictive power of MLMs appears far weaker than their proponents would say. Then in section 2.2, I will argue that Piantadosi's critique is based on a rather confused notion of explanation.

2.1 What do MLMs predict anyway?

There are two potential ways for a scientific model/theory to predict. Either they predict the outcomes of controlled experiments or they predict real-world future events. We can take quantum physics and meteorology respectively as canonical instances of sciences which meet those criteria. Quantum theory can calculate with a great deal of precision what the outcome of a well-designed experiment, though is utterly useless for predicting real-life events. Meteorology, on the other hand, deals with subject matter that cannot possibly be investigated in the lab, yet it is able to make accurate, albeit imprecise, predictions about the actual events.

With language as the subject matter, we can see what each type of predictiveness would consist in. The controlled experiments are precisely the judgement tasks and experiments of psycho- and neurolinguistics that populate the generative linguistics literature. For a theory/model of language to predict actual events, on the other hand, would be for it to predict actual utterances—to predict that the next sentence uttered by Jerry Patel of Port Arthur will be S, or that 5% of the sentences in tomorrow's *Financial Times* will be in the subjunctive mood, or the like. The closest we come to such predictions is to be found in the study of language variation and change, where broad trends can indeed be observed.

For all his touting of the predictivity of MLMs, Piandadosi (2023) does not cite any examples of it. Rather, he presents various outputs of ChatGPT and asks the reader to marvel at its "human-like behavior." To take an illustrative example, Piandadosi (2023, p. 16) prompts the chat-bot explain the significance of (2) to "[g]enerate ten other sentences like [it]."

(2) Colorless green ideas sleep furiously.

In response, the algorithm produces one partially meaningless sentence (3) and nine meaningful sentences such as (4) with the same syntactic structure as (2)

(3) Purple fluffy clouds dream wildly.

(4) Black shiny kangaroos hop playfully.

When it was pointed out that this was an instance of a failure of the model to be "human-like," Piandadosi (2023, p. 36) responded that

> it was strange to me that authors criticized these responses as showing the models do not exhibit 'human-like behavior' without even wondering whether people would produce meaningless bigrams from the same prompt.

His position, then, seems to be that every response of a chat-bot to a prompt is a prediction of how a person would respond to the same prompt. Such a prediction would be the easiest in the world to test, yet Piantadosi does not do so. This dereliction is understandable, though, as a moment's reflection is enough to realize that the odds of any human responding identically to the same prompt are infinitesimally small.

Baroni (2022), who is cited by Piandadosi (2023, p. 7), presents a more modest argument for MLM *qua* linguistic theories. The simplest statement of his claim is

> [W]e can think of a deep net architecture, before any language-specific training, as a general theory defining a space of possible grammars, and of the same network trained on data from a specific language as a grammar, that is, a computational system that, given an input utterance in a language, can predict whether the sequence is acceptable to an idealized speaker of the language. (p7)

There is little to object to in this claim. In fact, it echoes the Chomsky's (1965) definitions of descriptive and explanatory adequacy. So we can ask whether MLMs are descriptively adequate—do they correctly specify the class of possible grammars?[6]—and whether they are explanatorily adequate—do they define a function from a child's language input to a grammar of that language?[7]

I call this claim modest for two reasons. First, while descriptive and explanatory adequacy were considered ambitious goals in the mid-20th century, generative theory had moved beyond them by the turn of the 21st century (Chomsky, 2004). Current generative theorizing takes into account the fact that FLN is biologically instantiated and therefore was evolved and is constrained by laws of nature. Second, merely predicting "whether [a given] sequence is acceptable to an idealized speaker of the language" is a lower bar than that of even early generative theories, which were far more interested in semantics than is often thought. Indeed, the justification for transformations in grammar was the fact that sets of sentences were semantically linked to each other in systematic ways (Chomsky, 1957). Put another way, the grammars/models generated by MLMs aim to predict the outcome of any individual acceptability judgment, while transformational grammars and their successors aim to predict regular patterns in speaker intuitions and use.

A further prediction of MLMs, albeit an impicit one, involves intergenerational language transmission. If MLMs are good models of the human language faculty, then a trained MLM is a good model of a mature human language user. It follows from this that the data generated by MLMs will, in aggregate, be indistinguishable from actual human language data and therefore be suitable training data for subsequent generations of a model. Shumailov et al. (2024) test this prediction on the OPT-125m model (Zhang et al., 2022)[8] and find that, rather than reproducing the previous generation of model, this kind of "recursive training" very quickly leads to what they call "model collapse." That is, while a Gen 0 model—a model

[6] See section 1 for discussion of this.

[7] The amount of data required to train MLMs suggests a negative answer.

[8] From Meta Platforms, Inc., formerly Facebook Inc.

trained on real language data—recognizes real data well and generates human-like data, a Gen 9 model diverges noticeably from humans, recognizing far less actual language data and generating gibberish.

Shumailov et al. (2024) argue that this kind of model collapse is not a narrow property of the OPT-125m model that they test, but that all purely statistical models will collapse under recursive training. Model collapse, they argue, occurs primarily because any finite sample necessarily leaves out some data, with rare types of data being more likely to be left out. Therefore, some rare types of data are not represented in the proceeding datasets. They further argue that as recursive learning continues models "diverge arbitrarily far from the original one" and the variance in the model-generated data will go to zero.

It would seem then, that the predictions of MLMs hold only when limited to the very narrow domain of acceptability judgments—a domain that generative syntax has aimed to exceed since its inception. Once we go beyond that domain, the predictivity of MLMs begin to break down. Therefore, even if we accept Piandadosi's (2023) premises that scientific theories should be judged on their predictivity and MLM's are only improving, we can see that a declaration that MLMs can supplant generative theories of linguistics is, at best, premature.

2.2 Explanatory theories, Data, and Interpretation

One of the main critiques of MLMs *qua* theories is that, for all their apparent power, they are lack simplicity and explanatory power which are the hallmarks of scientific theories. This is most succinctly put by Rawski and Baumont (2023) who write "[e]xplanatory power, not predictive adequacy, forms the core of physics and ultimately all modern science." While Piandadosi (2023) does not dispute that MLMs are quite complex and not explanatory, he does dispute the centrality of those in science. He suggests, using the Large Hadron Collider and the James Webb Space Telescope as examples, Rawski and Baumont (2023) fail to understand the importance of data for modern physics[9], and refers to explanation as a "non-empirical consideration." Piantadosi wrongly frames explanation in opposition to data, but the former is inextricably linked to the latter.

At its simplest, the link between explanation and data is one of semantic entailment—theories are constructed to explain data. Or even at a more basic level, the verb *explain* is transitive, and its internal argument could reasonably be said to have the thematic role of "Data," much like the internal argument of *eat* has the thematic role of "food." So, to say that Rawski and Baumont's (2023) focus on explanation led them to forget about data is a bit like saying someone's focus on eating led them to ignore food—poetic, perhaps, but not a meaningful critique.

Explanation and data are linked in another way—a way that Piantadosi alludes to in his references to the Standard Model of quantum field theory and the Large Hadron Collider (LHC)—scientific explanation leads to the empirical discoveries. CERN did not build the LHC on a whim to gather more data as a good in itself, nor did physicists develop the Standard Model and its preceding theories merely as an excuse to generate more data. Rather, the pattern of scientific progress is as follows: First, theories are constructed to explain the available data. Second, scientists and mathematicians generate predictions by following the logic of those theories. Third, experimentalists design and run procedures which generate novel data to test those predictions. Finally, the resulting data is analyzed and, if some of the data cannot be explained by the current theory, the process starts again. For science, as it has been practiced at least since the time of Galileo, data is certainly important, but only insofar as it is relevant to the content of an explanatory theory.

[9]This is an odd response given that, while Piantadosi has no training as a physicist, Lucie Baumont is a cosmologist (*cf* Solnit, 2014 for a discussion of similar phenomena)

The reverse attitude—the prioritizing of data regardless of its relation to a scientific theory—has gained a great deal of prominence in contemporary society. This includes, but is not limited to, ostensibly scientific fields, but it does not follow from the standard scientific practice described above. Rather, the primacy of data preached by proponents of MLMs and required for the continued improvement of the models themselves (Kaplan et al., 2020) is more in line with the economic imperatives of capitalism described by Zuboff (2019).

A second argument Piandadosi (2023) makes against explanation is as follows. Generativists, according to Piandadosi (2023) demand that a theory be "*intuitively comprehensible* to us" (p35) yet this would require us to reject as unscientific the Copenhagen interpretation of quantum mechanics—summed up in Mermin's (1989/2016) aphorism "Shut up and calculate!" Therefore, "intuitive comprehensibility" cannot be a criterion for scientific theories. The implication here is that, quantum theory is inherently incomprehensible and, as a result physicists have abandoned explanation as a goal.

Piantadosi cites wave-particle duality as an aspect of quantum physics which is "impossible to intuit, but eminently scientific." It is likely true that no one would arrive at wave-particle duality merely from their day-to-day experience, nor would they arrive at it through pure reason. Yet the theories that predict wave-particle duality, like every scientific theory, depend on intuition in two ways. First, the development of wave-particle duality as a novel theory, required hypothesis (*i.e.*, guessing) which depends on some subconscious intuition. Second, the communication of the theory of wave-particle duality requires logical and mathematical argumentation, each step of which involves intuition.

To understand this last point, consider the statement (5) which is commonly used to introduce students to proofs.

(5) $0.999\cdots = 1$

Our immediate intuition says that this is false, but the reason this is an effective teaching tool is that we can convince ourselves of it using our mathematical intuition. Likewise the statement that a person standing still at the earth's equator is moving at a speed greater than 1500 km/h is intuitively false, yet Galilean mechanics leads precisely to that conclusion, and Galilean mechanics is an intuitively comprehensible theory. The similarity in these examples points to the flaw in Piantadosi's argument here: just because the results of a theory are counterintuitive doesn't mean the theory itself is intuitively incomprehensible.

In addition to this fallacy, Piantadosi conflates the notion of theory and interpretation which leads him to praise quantum mechanics and disparage generative linguistics for the same thing. Quantum theory is a set of equations which can be solved in various ways to make empirical predictions. The strangeness of the theory led to a variety of interpretations which attempt to give meanings to the equations. The most popular "interpretation," though, is the Copenhagen Interpretation which eschews meaning questions altogether—"Shut up and calculate!" Piantadosi approvingly cites this attitude as scientific, yet complains when generative theorists adopt a similar view—"There are *no* theories of how Principle C or A-chains or whatever are encoded in human brains, much less genomes." (p35) Though he is ostensibly asking for "theories" here, he is actually complaining about the lack of interpretation of the theory. The obvious answer to his complaint is "Shut up and calculate!"[10]—scientific theories don't need interpretations.

So, once we replace Piantadosi's muddled notion of explanation with a proper one, we can see that his critique falls apart. There is no conflict between explanation and data, provided data is not seen as an end unto itself. Furthermore, the generative search for explanation is fully in line with ordinary scientific practice.

[10] See Murphy (2020) for a recent neurological interpretation of minimalist syntax.

In arguing that MLMs supplant the generativist style of theorizing then, Piandadosi (2023) is actually putting forth a rather anti-scientific view. In his polemic, Piantadosi disparages explanation and understanding as scientific values despite the fact that they have been central to all of modern science. In their place, we are told that predictive power is what makes a theory/model scientific and asked to judge MLMs and generative theories by a very narrow domain of prediction seemingly defined by a tool from software engineering—a test suite (Gauthier et al., 2020). That is, he argues for replacing scientific aims and values with those of the technologist. This amounts to anti-science because the essence of science includes its aims and values.

3 Conclusion

Modern Language Models neither refute generative theories of grammar nor do they supplant the generative approach to language theorizing despite Piandadosi's (2023) claims to the contrary. They are not good models of the human language faculty, as they do not have the same limitations as we do, and they are not scientific theories as they offer no explanations of any aspect of nature. The simple reason for this is that MLMs are technologies, not theories—they are designed to serve some purpose for the corporations that built them, not to enhance our understanding of the world. Expecting them to serve as theories can only be the result of a category error.

Note this is not a criticism of MLMs any more than saying a car is ill-suited to cooking food is a criticism of cars. The criticism or praise of MLMs should come from judging them as technologies—from asking what MLMs do, who they do it to, and who they do it for (Doctorow, 2022). For instance, Marcolli et al. (2023) may be correct that MLMs will be useful in testing predictions of scientific theories of language, or perhaps their main uses will be less benign. Aside from what they are used for, we should also judge them for the effects they have on workers, the climate, inequality, and society as a whole. These concerns are beyond the scope of this chapter, though, and I leave them for everyone to consider for themselves.

References

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

Chomsky, N. (1957). *Syntactic structures*. Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Chomsky, N. (2004). Beyond explanatory adequacy. In A. Belletti (Ed.), *Structures and beyond* (pp. 104–131). Oxford University Press.

Chomsky, N. (2013). Problems of projection. *Lingua*, *130*, 33–49.

Chomsky, N., Roberts, I., & Watumull, J. (2023). Noam Chomsky: The false promise of ChatGPT. *The New York Times*.

Doctorow, C. (2022). Science fiction is a luddite literature. *Locus*, *732*, 26.

Galilei, G. (2001). *Dialogue concerning the two chief world systems* (S. Drake, Trans.). The Modern Library. (Original work published 1632)

Gauthier, J., Hu, J., Wilcox, E., Qian, P., & Levy, R. (2020). Syntaxgym: An online platform for targeted evaluation of language models. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 70–76.

Kallini, J., Papadimitriou, I., Futrell, R., Mahowald, K., & Potts, C. (2024). Mission: Impossible language models.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling laws for neural language models. https://arxiv.org/abs/2001.08361

Kodner, J., Payne, S., & Heinz, J. (2023). Why linguistics will thrive in the 21st century: A reply to piantadosi (2023). https://arxiv.org/pdf/2308.03228

Marcolli, M., Berwick, R. C., & Chomsky, N. (2023). Syntax-semantics interface: An algebraic model. https://arxiv.org/abs/2311.06189

Mermin, N. D. (2016). What's wrong with this pillow? In *Why quark rhymes with pork: And other scientific diversions* (pp. 23–28). Cambridge University Press. (Original work published 1989)

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

Moro, A., Greco, M., & Cappa, S. F. (2023). Large languages, impossible languages and human brains. *Cortex*, *167*, 82–85.

Murphy, E. (2020). *The oscillatory nature of language*. Cambridge University Press.

Musso, M., Moro, A., Glauchel, V., Rijntjes, M., Reichenbach, J., Büchel, C., & Weiller, C. (2003). Broca's area and the language instinct. *Nature Neuroscience*, *6*(7), 774–781.

Piandadosi, S. T. (2023). *Modern language models refute chomsky's apprach to language*. https://ling.auf.net/lingbuzz/007180

Rawski, J., & Baumont, L. (2023). *Modern language models refute nothing*. https://lingbuzz.net/lingbuzz/007203

Shumailov, Z., Ilia and Shumaylov, Zhao, Y., Papernot, R., Nicolas and Anderson, & Gal, Y. (2024). Ai models collapse when trained on recursively generated data. *Nature*, *631*(8022), 755–759. https://doi.org/10.1038/s41586-024-07566-y

Smith, N. V., & Tsimpli, I.-M. (1995). *The mind of a savant: Language learning and modularity*. Blackwell Publishing.

Solnit, R. (2014). *Men explain things to me and other essays*. Haymarket Books.

Tettamanti, M., Alkadhi, H., Moro, A., Perani, D., Kollias, S., & Weniger, D. (2002). Neural correlates for the acquisition of natural language syntax. *NeuroImage*, *17*(2), 700–709. https://doi.org/https://doi.org/10.1006/nimg.2002.1201

Zhang, S., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., Dewan, C., Diab, M., Li, X., Lin, X. V., Mihaylov, T., Ott, M., Shleifer, S., Shuster, K., Simig, D., Koura, P. S., Sridhar, A., Wang, T., & Zettlemoyer, L. (2022). Opt: Open pre-trained transformer language models. https://arxiv.org/abs/2205.01068

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.