



Nefdt, R. M. (2024). *The Philosophy of Theoretical Linguistics: A Contemporary Outlook*. Cambridge University Press. ISBN 9781009082853.

Metatheoretical Linguistics: A Philosopher's Guide

Konstantinos Voudouris^{1,2} , Niall Roe³ 

[1] Institute for Human-Centered AI, Helmholtz Zentrum Munich, Neuherberg, Germany. [2] Leverhulme Centre for the Future of Intelligence, University of Cambridge, Cambridge, United Kingdom. [3] Department of History and Philosophy of Science, University of Cambridge, Cambridge, United Kingdom.

Biolinguistics, 2024, Vol. 18, Article e15915, <https://doi.org/10.5964/bioling.15915>

Received: 2024-10-23 • Accepted: 2024-10-28 • Published (VoR): 2024-11-08

Handling Editor: Patrick C. Trettenbrein, Max Planck Institute for Human Cognitive and Brain Sciences & University of Göttingen, Germany

Corresponding Author: Konstantinos Voudouris, Institute for Human-Centered AI, Helmholtz Zentrum Munich, Ingolstädter Straße 1, D-85764, Neuherberg, Germany. E-mail: kv301@srcf.net

Abstract

In this article, we summarise and critically evaluate Ryan Nefdt's *The Philosophy of Theoretical Linguistics: A Contemporary Outlook* (2024). In this book, Nefdt brings the tools of philosophy to bear on contemporary linguistics, targeting perennial debates in syntax, semantics, pragmatics, phonology, and the evolution of language. In so doing, Nefdt sketches several tantalising paths for progress on these topics. Although some of Nefdt's arguments are underdeveloped, this book serves as a worthwhile introduction to the philosophy of linguistics.

Keywords

philosophy of science, linguistic theory

Ryan Nefdt's *The Philosophy of Theoretical Linguistics: A Contemporary Outlook* (2024) is a lucid guide and opinionated introduction to the under-explored, and *under-appreciated*, field that bears the same name. This monograph is part of a contemporary tradition in the philosophy of science that seeks to describe, clarify, and critically evaluate the practice of scientists, paying close attention to the theories and methodologies of a field and bringing philosophical tools and concepts to bear on them (Ankeny et al., 2011). While contemporary philosophers of science have written extensively on the practices of physicists, biologists, social, and cognitive scientists, engagement with theoretical linguistics has been relatively rare, particularly for sub-fields such as phonology, biolin-



guistics, and natural language processing. *The Philosophy of Theoretical Linguistics* seeks to put that right by sweeping across the breadth of modern linguistics, introducing key approaches, theories, and debates, and presenting insightful analyses that draw on the tools of contemporary philosophy.

The substantive chapters of Nefdt's book follow a roughly tripartite structure. First, he targets a sub-domain (in the case of Chapters 3–7 on syntax, semantics, pragmatics, phonology, and natural language processing) or debate (in the case of what counts as a possible language in Chapter 2 or how language might have evolved in Chapter 8). Second, he introduces a toolkit of philosophical concepts, such as possible world semantics (Chapter 2) or the role of explanation and prediction in science (Chapter 7), and theoretical approaches from other scientific fields, such as action theory (Chapter 6) and systems biology (Chapter 8). Third, he uses the toolkit to clarify key properties of his linguistic quarry and plot a path for progress. Theories, debates, and conceptual machinery are deftly woven together, showcasing Nefdt's command of both theoretical linguistics and contemporary analytic philosophy.

The fundamental accomplishment of this book is to build bridges, between linguistics and philosophy, between linguistics and other sciences, and between linguistic theories themselves. Our fundamental concern is that Nefdt's bridges cannot yet be traversed. From possible world accounts of infinitude to action theory for phonology, Nefdt sketches several paths towards applying philosophical concepts and analyses to methodological questions in linguistics, but they remain sketches. While a philosophy of science for linguistics is much needed, *The Philosophy of Theoretical Linguistics* reads as a first foray into what ought to be a rich and historied field of philosophical research. It is hampered by its impressive diversity: Nefdt does not give himself the room to fully convince the sceptical reader about how they might profit from the bridges that he starts to build. Each chapter could and should serve as the basis for its own book, and indeed, such books do exist (e.g., Hale & Reiss, 2008; Hinzen & Sheehan, 2013; Kempson et al., 2012; Ludlow, 2011). This book is a worthwhile starting point, but it is just that.

In this article, we summarise and critically evaluate Nefdt's book. First, we turn to Chapters 1, 3, 4 and 5, which seek to build bridges between linguistic theories that have traditionally been seen as incompatible. Next, we turn to Chapters 6 and 8, which seek to build bridges between topics and debates in theoretical linguistics and theories from elsewhere in the sciences. Then, we examine Chapters 2 and 7 which use tools from analytic philosophy to clarify and make progress on debates in linguistics. In the penultimate section, we outline one problem for the thesis that Nefdt develops across Chapters 2, 6, 7, and 8, namely, that we should embrace broader conceptions of language so that we can employ a more diverse toolkit of methodologies to study it. The final section summarises our review.

Bridges Between Linguistic Theories

Nefdt sketches out the scope of the *Philosophy of Theoretical Linguistics* in Chapter 1, outlining what we can take to be the purview of *theoretical linguistics*, its relation to the cognitive and social sciences, and the key methodological tools that theoretical linguists use, including grammaticality judgments and formal theoretical descriptions. The central claim is that, for all the diversity in approach, method, and metaphysical commitment, theoretical linguists are united by a commitment to a small set of key questions (p. 3):

1. a) What is Language? → b) What is a language?
2. How do we acquire languages?
3. How is linguistic communication possible?
4. How did language evolve?

To make progress on Questions 2–4, theoretical linguists stake their claim on certain answers to Question 1, thus making the two flavours of that question of ultimate importance to the field. Nefdt's insight here, in laying out the unifying questions of the field, is an encouraging reminder that researchers of all stripes can find unity under the banner of these questions, despite the often vitriolic debates that have plagued contemporary linguistics.

Indeed, Nefdt's tactic of identifying a key question that *all* researchers are seeking to answer, is something that he takes forward to unifying diverse research programmes in syntax and semantics in Chapters 3 and 4. In Chapter 3, *Syntactic Metatheory*, Nefdt posits that all theories of syntax aim to characterise three basic properties of human languages, and more specifically, human syntaxes:

1. That there are *rules* that govern the combination of more complex units from more simple units, i.e., syntax is *rule-bound*.
2. That the rules for combining units are represented in a way that is irreducible to other kinds of linguistic representations (such as semantic representations), i.e., syntax is *autonomous*.
3. That (at least some of) the rules for combining units are recursive, i.e., syntax exhibits *recursion*.

Nefdt then shows that modern theories of syntax emerge from enriching those basic properties. For instance, Minimalism (e.g., Chomsky, 1992, 1995) claims that syntactic rules create hierarchical structures by composing *constituents*, leading to phrasal structure. In contrast, Dependency Grammar (e.g., Osborne, 2014) claims that syntactic rules create hierarchical structures by specifying asymmetric dependency relations between primitive units. Similar considerations apply to model-theoretic theories such as Head-Driven Phrase Structure Grammar (HPSG; Pollard & Sag, 1994). These theories are all committed to the claim that syntax is rule-bound, but they differ in what those rules represent.

To strengthen his claim about the basic conception of human syntax, Nefdt traverses as far as he can from formal theories, to those reactionary approaches that seek to reject the central tenets of mainstream generative syntax. He presents Croft's (2001) Radical Construction Grammar (RCG), which purports to reject formal syntactic structure altogether, replacing it with language-specific idiomatic, non-decomposable constructions. Even here, Nefdt argues, we can see a commitment to the basic conception of human syntax. He argues that constructions such as *X is more Adj than Y* clearly invoke some rule-based structure—only certain units can fill the gaps and these are specified by rules of some kind, even if they aren't of the inviolable, universal type that are familiar to generative syntacticians. Moreover, he demonstrates that constructions can also be recursive, being embedded in themselves with no clear termination criterion beyond memory and time limitations. Finally, Nefdt points to claims from RCG practitioners that constructions can be purely formal templates, with no semantic value (p. 76), suggesting that syntactic representations are at least partially autonomous from semantic ones. Despite the diversity of theories of syntax that exist, including those that aim to explicitly reject the assumptions of others, Nefdt claims to have found the core properties of human syntax on which bridge all theories of it.

In Chapter 4, Nefdt continues this approach, turning to the field of semantics. He advances Dowty's (2007) claim that a fundamental property of human language, and specifically, the human semantic system, is that it is compositional, but that specifying what that compositionality amounts to is up for debate. Nefdt shows how contemporary semantic theories are consonant with this claim, providing different answers to the question of what compositionality is, starting with Montagovian semantics and dynamic theories such as Discourse Representation Theory and finishing with lexical semantics, and super-semantic theories that seek to characterise non-linguistic meaning (such as is conveyed by gesture).

Chapter 5 starts to demonstrate, in our view, the limits of this approach, or at least, Nefdt's struggle to adequately define the key attributes of his target research programmes. In this chapter, Nefdt addresses the field of pragmatics. His main aim is to characterise how the scope of pragmatics differs from semantics, identifying differing views on the nature of *context*, and how interlocutors interact with it, as the critical arbiter. He outlines three philosophical views that structure the discussion. He starts with Grice (e.g., Grice, 1989), who views context as the static background against which speakers make inferences about *what is meant* (what the speaker intends to communicate) based on *what is said* (the literal meaning of the utterance). The latter is to be explained by semantics, while the former is to be explained by pragmatics. Under the Gricean view, speakers use the conjunction of literal semantic meaning, the context in which the utterance takes place, and the assumption that interlocutors are cooperating with each other, to infer what is meant. How speakers perform these inferences is then described by the mechanics of various Gricean and neo-Gricean theories of pragmatics.

Nefdt contrasts the Gricean approach to pragmatics with the perspective of Stalnaker (1978), who argues that pragmatics is tasked with modelling how context changes over the course of a conversation. For Stalnaker, context is a dynamic set of propositions that interlocutors co-create, with the goal of producing utterances that reduce the set of possible worlds that are consonant with the set of propositions (converging on only *this* world in the infinite conversational limit). Utterances have a semantic content (corresponding to Grice's *what is said*) but also a pragmatic content, inferred on the assumption that interlocutors will produce utterances that change the context in the direction of reducing the set of possible worlds compatible with it. The third philosophical perspective, from Lewis (1979), who views pragmatics as tasked with modelling how interlocutors keep conversational score, which is the set of presuppositions and common ground shared between the speakers to which each offers new information, about, for instance, what indexicals refer to. In phrasing it this way, Lewis views pragmatics as modelling conversations game-theoretically, beyond the remit of classical semantic analysis.

Nefdt then attempts to put these contrasting philosophical perspectives to work for understanding the debated semantics-pragmatics distinction, of which he argues there are three conceptions. The indexicalist view sees pragmatics as an extension of semantics, dealing with the parts of language that require context for their meaning to be fixed, namely, indexicals. The distinction is kept clean: “[s]emantics studies something about non-indexical languages and pragmatics studies something about indexical ones” (Szabó, 2009, p. 369; cited in Nefdt, 2024, p. 120). Nefdt argues that this conception most neatly fits with Lewis's conception of pragmatics (p. 119) as modelling how to keep conversational score, but this relation is never fully fleshed out. In contrast, the cognitivist view sees semantics and pragmatics as being cognitively distinct processes, controlled by different parts of the brain. The contemporary version of this view is Relevance Theory (Sperber & Wilson, 1986), which unsurprisingly is linked to the Gricean picture of pragmatics.¹ The final take on the semantics-pragmatics distinction is that of Horn and Kecskes (2013), referred to as *Socio-Cultural Interactional Pragmatics* (SCIP). This view sees pragmatics as encompassing much more than the preceding views, aiming to model social and cultural norms and speaker meaning across whole discourses (not just single utterances), as well as implicatures, presuppositions, and other standard fare of pragmatic theory (Levinson, 1983). Pragmatics is, thus, a social phenomenon in a way that semantics is not, although the precise implications for the semantics-pragmatics distinction is lost in the discussion here. SCIP is connected to Stalnaker's view of context as the holistic set of propositions and the possible worlds with which they are compatible. While these examples give a wide view over the field of pragmatics, it is not

1) Unsurprising since Relevance Theory is unashamedly neo-Gricean, subsuming all pragmatic inference under a version of Grice's maxim of relevance.

clear what the argument is supposed to be. Nefdt misses the opportunity to flesh out his fascinating claim that modern theories of pragmatics differ in what they take *context* to be, and jumps straight to the well-hammered debate over the semantics-pragmatics distinction. The theories he presents certainly differ in what they view to be the remit of pragmatics as opposed to semantics, but connecting them to the philosophical views of Lewis, Grice, and Stalnaker does little to help readers understand what is at stake in these disagreements. The connection to the role and nature of context is offered is pregnant with possibility here but never borne out in the text.

This situation deteriorates further in the final part of Chapter 5, in which Nefdt presents some lesser known theories of pragmatics, including an Optimality Theoretic approach, a Bayesian analysis, and a game-theoretic model. The mastery of this diversity of theory is no doubt impressive, but Nefdt is forced to conclude that ‘[i]n reality, each [pragmatic] theory incorporates various aspects of the philosophical conceptions, with a bias towards cognitivist approaches’ (p. 135), leaving the reader unsure of what to take away from the discussion. Unlike Chapters 3 and 4 on syntax and semantics, the core properties of the three philosophical views are difficult to discern and discriminate. It is difficult to leverage them for analysing what contemporary pragmatic practice is all about and understanding what bridges these disparate theories.

Bridges Between Scientific Theories

Nefdt takes an alternative approach in Chapters 6 and 8. Instead of analysing the core properties of a sub-field, clarifying what is at stake therein, he attempts to build bridges between linguistics and other scientific fields. With a philosopher’s eye, Nefdt identifies commonalities in the explananda of different sciences and plots a path for progress on linguistic questions by integrating them. The implication here seems to be that insights from other sciences will offer new perspectives on perennial debates in linguistics. Chapter 6 focuses on phonological theory, a domain that has received little attention in the philosophy of science (although see Carr, 2012; Hale & Reiss, 2008). After introducing the basic concepts of phonological theory as they apply to both spoken and signed language, Nefdt suggests that theories of motor action would be a fruitful way to model phonological phenomena. Chapter 8 focuses on the evolution of language, arguing for an alternative to contemporary biolinguistics accounts (see, e.g., Berwick & Chomsky, 2016; Boeckx, 2013), namely, that “natural language is a complex system and its emergence is likely to have been prompted by multiple interacting factors.” To advance this proposal, which Nefdt calls the Maximalist Program, he suggests drawing on systems biology and complex systems theory. This has the positive auxiliary consequence of grounding theories about the evolution of language in the biological sciences.

The similarities between the evolution of language and the structure of biological networks, as well as between phonology and action theory, are well motivated in Nefdt’s

discussions. However, they lack a fully-fledged account of what these novel proposals can do for theoretical linguistics. Of course, such accounts would demand far more than the 20 pages or so that Nefdt has given himself, but the reader is left wanting. Moreover, it is not entirely clear that Nefdt's proposals are completely novel. He summarises empirical and theoretical work that build action-theoretic models of phonology (e.g., Schwartz et al., 2007), and even states that “[t]here are too many theories and models to consider here” (p. 155). It is unclear whether Nefdt is advocating for something that differs in degree or kind from those approaches. Similarly, a systems approach to the study of language is the topic of Kretzschmar's (2015) book *Language and Complex Systems* and Kirby's (1999) book *Function, Selection, and Innateness: The Emergence of Language Universals*, both of which are cited by Nefdt. The novelty of Nefdt's proposals is not substantiated in the text. These chapters thus serve as useful introductions for the linguistically untrained philosopher to the fields of phonology and language evolution, but it is unclear what these chapters offer the practising linguist.

Bridges Between Linguistics and Philosophy

Chapters 2 and 7 contain some of the book's most direct engagement with philosophical concepts. Chapter 2 focuses on the distinction between *Language* and *languages*, which Nefdt identifies as a special case of a well-worn question: how to define a *general* through studies of *particulars* (Ramsey, 1931; Strawson, 1954). In Chapter 7, Nefdt approaches the problem of reconciling contemporary linguistic theory with the results of modern machine learning by examining how linguists treat explanation and prediction as epistemic values. In both chapters, Nefdt seeks to clarify what is at stake in contemporary debates by using the concepts of analytic philosophy.

Chapter 2 investigates what constitutes a possible human language. Nefdt begins by noting that a merely empirical approach to this question will never reveal more than the tendencies of actual languages, failing to adequately describe the set of languages that *could* exist. While formal approaches such as those employed by generative linguists are superior, they too have their shortcomings. He argues that it can be difficult to tell whether a property of a formal model belongs to (just) the model, or to the target being modelled. Recursion may be merely a feature of our theory, not of the world, for example. While this argument is gestured at in the book, it is only developed in a separate paper (Nefdt, 2019). If this argument stands, then formal models also do not help us delimit a possibility space for Language, but only a local possibility space for Language relative to a model.

Nefdt substantiates this argument with a critique of Andrea Moro's (2016) work on impossible languages, which exemplifies this local possibility space. Moro defines Language formally as any combinatorial system exhibiting Merge, and hypothesises that hierarchically structured systems (“possible languages”) will be learned differently to lin-

early structured ones (“impossible languages”). Moro’s results indicate that hierarchical languages use different areas of the brain to linear ones.² Nefdt’s criticism of Moro’s work begins by noting that the properties in question (i.e., recursion) may be properties of the model, not of the language. Though, as mentioned, the details of this critique are left to a separate paper (Nefdt, 2019). Nevertheless, Nefdt is right to point out that it is not clear exactly which property accounts for Moro’s experimental results. It could be hierarchy, structure dependence, recursion, or the fact that there are some infinitely productive rules (p. 35). Even if Moro staked a claim here, Nefdt suggests that these concepts interrelate and are therefore hard to pick apart as difference makers. Nefdt also suggests that Moro’s results are not satisfying because the subjects did indeed learn the impossible languages, undermining the claim that they are truly impossible. This critique, we think, puts the cart before the horse. As Nefdt initially stresses, only once we have defined Language are we in a position to describe how such a thing is learned or used for communication. Moro’s claim was never that certain systems were unlearnable, but that *if* Language is defined by the presence of the Merge operation, *then* we should expect certain differences in how people learn hierarchically vs linearly structured sequences, evidence of which was found. Nefdt’s main gripe with Moro’s approach appears to be its narrowness. Moro gives us no way to talk about whether some hypothetical language would be more or less possible. However, this does not appear to have ever been Moro’s project.

In search of a more refined notion of possibility, Nefdt marshals modal logic, possible world semantics (Berto & Jago, 2019), and Lewis’s (1983) notion of “naturalness”. Here, Nefdt takes the natural properties of language to be learnability and usability. Every *actual* natural language has these features, and this characterization excludes formal languages, and languages with inconsistent sets of properties, from being possible natural languages. Furthermore, he posits that possible languages exist on a spectrum. Languages on this spectrum are deemed more or less possible depending on how modally “close” they are to actual language—that is, by the number of modifications one would have to make to existing linguistic theory (or related fields) to accommodate them. For Nefdt, “a possible language is a [...] *way language could have been*” (p. 43).

While the application of modal logic to this area of linguistics is innovative, it is unclear how it moves the needle on the question of possible languages. Nefdt rightly acknowledges that the possibility space defined by his account depends on how one characterises language at the outset (p. 47). This follows from his first chapter, where questions related to how language is used, learned, etc., all depend on how language itself is characterised. But in defining language in terms of usability and learnability,

2) This result is consonant with other work that Nefdt does not describe. For instance, Tsimpli and Smith (1995) presented a language savant, Christopher, with ‘possible’ and ‘impossible’ languages as defined by the formal constraints of generative grammar, and found that he could only learn the former.

Nefdt opens himself to the very critique he levelled against Moro. We see no reason why usability and learnability are any less theory-laden, or any more “natural”, as concepts than recursion, except that Nefdt spends significantly less time defining them than Minimalists have spent defining Merge! Ultimately, Nefdt’s proposal stands as an equal to Moro’s, not a superior. It offers an alternative that one *could* take if one preferred, but an argument for why one *should* take it is lacking.

Chapter 7 places linguistics within a larger narrative in the history and philosophy of science. Nefdt argues that linguistics has been unduly focused on explanation, resulting in the unjust neglect of prediction as a scientific virtue. This is due to an overcorrection away from positivistic and behaviouristic ideals. Both of these approaches foregrounded prediction and downplayed explanation in an attempt to rid science of apparently unscientific metaphysics. Under these views, any theory could explain, but the best scientific theories could also predict. For example, pigeon behaviour is trivially explained by describing them as *hungry*, but true value comes from accurately predicting behaviour as a function of satiety. The general trend away from positivism was accelerated in linguistics under the ardent influence of Chomsky (1959, 1965), whose research program emphasises explanation. However, Nefdt provides a case study in which prediction is emphasised over explanation, namely, theories of Negative Polarity Items (NPIs; e.g., “X didn’t ever see” vs “X ever* saw”). Nefdt argues that theories of NPIs are deemed successful insofar as they can predict environments in which NPIs occur. Nefdt takes this case study to demonstrate that prediction and explanation are essentially connected in linguistics (p. 163). We wish to note, however, that Chomsky and Chomskyans are not against admitting that prediction is important for science. Rather, one must contrast explanations with mere descriptions, or predictions that arise out of generalisations of such descriptions (e.g., Chomsky, 2023).

Making more room for prediction allows us to make use of methods ignored by narrower approaches limited by their characterization of language and focus on explanation. Such limited approaches, Nefdt argues, utilise methods unfit for the “noisy messy data” of actual linguistic corpora—something to which stochastic, continuous methods such as machine learning are well suited (p. 168). Much like how modal logics allow for a gradation of how *possible* a language is, stochastic, continuous methods allow us to model grammaticality on a sliding scale. Nefdt summarises the main issues with such stochastic methods (Pullum, 2009), but does not engage with them. Instead, he gives a sketch of one way we might *stochasticise* grammar. He does not defend this approach either, but instead wants to show that it clearly belongs in theoretical linguistics (p. 171). He does so with a detailed manifesto for the inclusion of machine learning approaches that extract structure from large corpora using statistical methods and little in the way of formal linguistic theory. In particular, Nefdt presents evidence of such approaches succeeding in modelling syntactic agreement and filler gap dependencies (although they continue to struggle with centre embedding). Further, he reviews literature on how

machine learning may call into doubt poverty of the stimulus arguments. Based on Lappin and Shieber (2007), he shows that deep neural networks are able to successfully learn languages remarkably well with a weak set of starting biases. This suggests, it is argued, that language need not arise from a rich, innate system of organisation, since it can be learnt from the latent structure of large input data.

Nefdt argues that machine learning presents an opportunity for theoretical linguistics because of what it can offer for prediction, and that it has been marginalised for that very reason. Nefdt diagnoses the preference for explanation over prediction as being caused by the so-called *Galilean style*—an approach to scientific inquiry defended by Chomsky whereby scientists abstract away from the inevitable messiness of the real world to investigate some idealised version of it. This style requires ignoring aspects of our experience that disagree with the idealisation. Such an approach is not overly interested in empirical descriptions—as every idealisation is descriptively false from the outset. Instead it focuses on explanation: which idealisation can account for the most relevant data? In critiquing this view, Nefdt questions Chomsky’s historical and descriptive claims, suggesting that science is not as methodologically monist as Chomsky believes. He cites Strevens (2020) “Iron Rule” as a counterexample that emphasises empirical considerations in the development of science.

However, we believe Nefdt’s characterisation, pitting theory *against* data, may underplay a crucial aspect of the Galilean style. Namely, that data only become data *of interest*—potential evidence—in light of a particular research question or theoretical framework.

It is a mistake to argue, as many do, that by adopting this point of view [the Galilean style] one is disregarding data. Data that remain unexplained by some coherent theory will continue to be described in whatever descriptive scheme one chooses, but will simply not be considered very important for the moment.

(Chomsky, 1980, pp. 11–12)

A geometry student might justifiably refuse to use a protractor to answer their homework questions. This is not because they deny that triangles can or should be empirically measured, but because their interest is in a different, formal aspect of triangles. Every scientific endeavour is similarly limited. For example, in Nefdt’s argument for embracing machine learning methods, he rightly notes that doing so requires ignoring certain empirical discrepancies, in the hopes they could be modelled later. Namely, unlike Deep Neural Networks (DNNs), “children don’t learn language by scanning the Wall Street Journal” (pp. 178–179).

In our view, linguists should embrace DNNs to the extent that DNNs are shown to be relevant to their project. Here, Chomsky is extreme, suggesting that DNNs are artefacts of clever engineering, but offer no scientific insights into the structure and function of

human language (Chomsky, 2023, p. 363). Justifiably or not, he has limited his focus to a posited system in our brains, not just any system that can produce a similar output. Moro et al. (2023) argue that DNNs are not suitable models for human language because they learn possible and impossible languages equally well—a consideration obviously core to their research program. Others will likely be more easily convinced that DNN performance is relevant to their linguistic interests. But this is not because they are being convinced to embrace the primacy of data over theory, but because they are being shown how given data is relevant to their project. Nefdt takes some steps in trying to convince the generativist that DNNs are relevant to them when discussing the poverty of the stimulus arguments. But primarily he appears to be arguing that linguists ought to adopt a different conception of language that would let them make use of the resources offered by machine learning.

A Missed Step

Nefdt's eagerness to redefine 'language' to make better use of machine learning gets to what we see as a key problem with the account of pluralism motivated by discussions in Chapters 2, 6, 7 and 8. In each, Nefdt argues for a conception of language that broadens our attention, rightly suggesting that this could be of some benefit to linguistics as a whole. The logic of "linguistic possibility" provides a global account of possibility rather than Moro's merely local version; viewing phonological phenomena as being actions opens the door to integrating phonology with the broader study of motor systems; a broad view of language allows us to use machine learning to test linguistic hypotheses; and the Maximalist Program embraces an account of language as a complex system, allowing us to use more of our evolutionary biology toolkit. These broader conceptions of language have much to recommend them, and Nefdt is right to note that linguists need not commit themselves to any particular narrow definition. He is also right to point out that alternative conceptions of language allow us to make use of more interdisciplinary tools, an upshot he terms "methodological pluralism". And he may well be right in his claim that this is where the "juiciest fruit" is to be found (p. 206). Nevertheless, our worry is that this approach is focused on developing the wrong kind of diversity; it is focused on creating a dish with a wide range of flavours, rather than a menu with a variety of dishes.

In each of these chapters Nefdt argues that the broader conception of language should be adopted over a narrower one, in part because it is broader and so allows for a greater variety of investigations. But such practical considerations cannot be allowed to dictate how scientists conceive of language. It is of course right to say that we cannot pursue an end without some possible means, but the best end is not necessarily the one with the most means available to it. If taken the wrong way, Nefdt's arguments from methodological pluralism would find us digging for treasure at the hardware store

because it houses the largest variety of shovels. We believe that the pluralism responsible for the richness of linguistics comes—at least in part—from a variety of answers to the unifying question: “what is language?” many of which beget limited, incompatible research programs. Based on the first chapter, we imagine Nefdt would agree with this point—this *question* is what unifies a diverse field. But in Chapters 2, 6, 7 and 8 he appears instead to be arguing for a single answer to that primary question. In each he argues that we ought to conceive of language as whatever lets us approach the subsequent questions with the greatest plurality of tools.³

Methodological pluralism of this kind might instead be called kaleidoscopic monism—a plea to characterise language as a single, multifaceted entity that can be approached in a number of ways. That is a promising characterization, but does not justify excluding others, such as the generativist approaches so often used as Nefdt's foil. But Nefdt's intentions here are unclear. If his broad characterization is meant to replace narrower ones, then Chapters 2, 6, 7 and 8 are at odds with Chapter 1. Indeed, he concludes Chapter 7 as if he has nearly completed his task of refuting narrow conceptions of linguistics (p. 181). However, if his broad approach is meant to sit alongside narrower ones as a single dish on an eclectic menu, a view we would embrace, then Chapters 2, 6, 7 and 8 lose much of their argumentative force. This is because Nefdt's broad conception doesn't promise a better source of answers than those it would replace. Rather, in providing a different definition of the object of inquiry it shifts the focus to a different set of problems. Nefdt provides many good arguments critical of the generativist program, but they are far from decisive. The above is not meant to suggest that the Chomskyan approach is preferable to Nefdt's, but instead to suggest that the most productive pluralism makes room for both. By emphasising a singular—though rich—definition of language, Nefdt risks sidelining a diversity of narrower approaches.

A Bridge Not Taken?

In arguing for methodological pluralism, Nefdt sketches a number of interdisciplinary connections, through which the book delivers on its introductory promise of bringing diverse philosophical and scientific perspectives to bear on various aspects of linguistics. Thus, the book also is likely to realise its concluding hope: to start a number of interdisciplinary conversations. However, the book falls short of delivering a philosophy of theoretical linguistics.

Nefdt's ambition is praiseworthy, but puts him in a difficult position. Part of his point is to showcase linguistics as a diverse and developed science, consisting of various subfields, methodologies, and guiding principles. Nefdt shows that linguistics is a ripe

3) It appears as though Nefdt (2023) develops this view in a book-length treatment, but reviewing this is beyond our scope.

(and unjustly neglected) territory for philosophers of science. His difficulty is that in order to demonstrate the richness of linguistics, he first must walk his readers through the landscape, including the subfields covered in this book as well as the various related philosophical and scientific domains he ably introduces and applies.

The introductory chapter hints at a way of understanding how these diverse strands could be braided into a cable, anchored in fundamental questions. To braid this cable would indeed require first being introduced to each strand, but the philosophical work would be in showing how each relates to the others, and how those relations contribute to or hinder the purpose the cable is meant to serve. Nefdt's hierarchy of questions begins this project in the first chapter, but it is taken no further. And while the book describes each strand, the forest is somewhat lost to the trees. Forced to cover so much ground in limited space, Nefdt must traverse it quickly. We wish he had spent less time on certain technical summaries, impressive though they were, and instead devoted more words to developing the bigger philosophical picture.

Because of Nefdt's able handling of the gamut of theoretical linguistics, this book serves as an accessible introduction to linguistics for philosophers. This introduction is well supplemented by the helpful suggestions for further reading at the end of each chapter. If one's philosophical work happens to be in one of the domains Nefdt draws upon, that particular chapter will be of special interest. The book will perhaps feel disappointing to the philosopher of science who wants a general account of linguistics, though it should whet their appetite. Meanwhile, the philosophically inclined linguist will come away with a number of tantalising suggestions for how to better understand the debates of their field, but little in which to sink their teeth. Indeed, the philosophically disinclined linguist would be forgiven for leaving this book without a taste of what philosophy of science can do for clarifying methodological practice—without a taste of how a philosophy of theoretical linguistics can help *them*.

Our criticism may appear unfair, as Nefdt only presumes to start interdisciplinary conversations, but the title and first chapter of the book hint at a road, or a bridge, not taken. Any dissatisfaction we have expressed about this work comes out of how important we think this bridge is, and how much we wanted Nefdt to take us over it. *The Philosophy of Theoretical Linguistics* certainly leaves us eager for more, and we hope that it can recruit others to help mine this rich vein.

Funding: KV was financially supported by the Templeton World Charity Foundation (TWCF-2020-20539) and the Helmholtz Zentrum, Munich. NR was financially supported by a Canadian SSHRC Doctoral Fellowship, a Cambridge Trust Scholarship, a Rausing, Williamson and Lipton Studentship, and a J. William E. Mingo Memorial Scholarship.

Acknowledgments: Many thanks to members of Marta Halina's HPS PhD Reading Group at the University of Cambridge for their comments on a previous draft.

Competing Interests: The authors have declared that no competing interests exist.

References

- Ankeny, R., Chang, H., Boumans, M., & Boon, M. (2011). Introduction: Philosophy of science in practice. *European Journal for Philosophy of Science*, 1(3), Article 303.
<https://doi.org/10.1007/s13194-011-0036-4>
- Berto, F., & Jago, M. (2019). *Impossible worlds*. Oxford University Press.
- Berwick, R., & Chomsky, N. (2016). *Why only us? Language and evolution*. MIT Press.
- Boeckx, C. (2013). Biolinguistics: Forays into human cognitive biology. *Journal of Anthropological Sciences*, 91, 1–28.
- Carr, P. (2012). The philosophy of phonology. In R. M. Kempson, T. Fernando & N. Asher (Eds.), *Philosophy of linguistics* (pp. 403–444). Elsevier.
- Chomsky, N. (1959). Review of Skinner's verbal behavior. *Language*, 35(1), 26–58.
<https://doi.org/10.2307/411334>
- Chomsky, N. (1965). *Aspects of a theory of syntax*. MIT Press.
- Chomsky, N. (1980). *Rules and representations*. Columbia University Press.
- Chomsky, N. (1992). A minimalist program for linguistic theory. *MIT Occasional Papers in Linguistics*, 1, 1–52.
- Chomsky, N. (1995). *The minimalist program*. MIT Press.
- Chomsky, N. (2023). Genuine explanation and the strong minimalist thesis. *Cognitive Semantics*, 8(3), 347–365.
- Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press.
- Dowty, D. (2007). Compositionality as an empirical problem. In C. Barker & P. Jacobson (Eds.), *Direct compositionality* (pp. 14–23). Oxford University Press.
- Grice, H. P. (1989). *Studies in the way of words*. Harvard University Press.
- Hale, M., & Reiss, C. (2008). *The phonological enterprise*. Oxford University Press.
- Hinzen, W., & Sheehan, M. (2013). *The philosophy of universal grammar*. Oxford University Press.
- Horn, L., & Kecskes, I. (2013). Pragmatics, discourse, and cognition. In S. Anderson, J. Moeschler, & F. Reboul (Eds.), *The language-cognition interface* (pp. 353–375). Librairie Droz.
- Kempson, R. M., Fernando, T., & Asher, N. (2012). *Philosophy of linguistics*. Elsevier.

- Kirby, S. (1999). *Function, selection, and innateness: The emergence of language universals*. Oxford University Press.
- Kretzschmar, W. (2015). *Language and complex systems*. Cambridge University Press.
- Lappin, S., & Shieber, S. M. (2007). Machine learning theory and practice as a source of insight into universal grammar. *Journal of Linguistics*, 43(2), 393–427.
<https://doi.org/10.1017/S0022226707004628>
- Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.
- Lewis, D. (1979). Scorekeeping in a language game. *Philosophical Papers*, 1, 233–249.
- Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61(4), 343–377. <https://doi.org/10.1080/00048408312341131>
- Ludlow, P. (2011). *The philosophy of generative linguistics* (1st ed.). Oxford University Press.
- Moro, A. (2016). *Impossible languages*. MIT 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>
- Nefdt, R. M. (2019). Infinity and the foundations of linguistics. *Synthese*, 196(5), 1671–1711.
<https://doi.org/10.1007/s11229-017-1574-x>
- Nefdt, R. M. (2023). *Language, science, and structure: A journey into the philosophy of linguistics*. Oxford University Press.
- Nefdt, R. M. (2024). *The philosophy of theoretical linguistics: A contemporary outlook*. Cambridge University Press.
- Osborne, T. (2014). Dependency grammar. In A. Carnie, Y. Sato, & D. Siddiqi (Eds.), *The Routledge handbook of syntax* (pp. 604–627). Routledge.
- Pollard, C., & Sag, I. (1994). *Head-driven phrase structure grammar*. University of Chicago Press.
- Pullum, G. K. (2009). Computational linguistics and generative linguistics: The triumph of hope over experience. *Proceedings of the EACL 2009 Workshop on the interaction between linguistics and computational linguistics: virtuous, vicious or vacuous?*, 12–21.
- Ramsey, F. P. (1931). *Foundations of mathematics and other logical essays*. Routledge.
- Schwartz, J.-L., Boë, L.-J., & Abry, C. (2007). Linking the Dispersion-Focalization Theory (DFT) and the Maximum Utilization of the Available Distinctive Features (MUAF) principle in a Perception-for-Action Control Theory (PACT). In M. J. Solé, P. S. Beddor, & M. Ohala (Eds.), *Experimental approaches to phonology* (pp. 104–124). Oxford University Press.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Blackwell.
- Stalnaker, R. (1978). Assertion. *Syntax and Semantics*, 9, 315–332.
- Strawson, P. F. (1954). Particular and general. *Proceedings of the Aristotelian Society*, 54(1), 233–260.
<https://doi.org/10.1093/aristotelian/54.1.233>
- Strevens, M. (2020). *The knowledge machine: How irrationality created modern science*. Liveright Publishing.
- Szabó, Z. G. (2009). The distinction between semantics and pragmatics. In E. Lepore & B. Smith (Eds.), *The Oxford handbook of philosophy of language* (pp. 361–390). Oxford University Press.
- Tsimpili, I., & Smith, N. (1995). *The mind of a savant: Language learning and modularity*. Blackwell.