



Simplicity of what? A case study from generative linguistics

Giulia Terzian¹ · María Inés Corbalán^{2,3}

Received: 9 September 2019 / Accepted: 31 March 2020 / Published online: 22 April 2020
© The Author(s) 2020

Abstract

The Minimalist Program in generative linguistics is predicated on the idea that simplicity is a defining property of the human language faculty, on the one hand; on the other, a central aim of linguistic theorising. Worryingly, however, justifications for either claim are hard to come by in the literature. We sketch a proposal that would allow for both shortcomings to be addressed, and that furthermore honours the program's declared commitment to naturalism. We begin by teasing apart and clarifying the different conceptions of simplicity underlying generative inquiry, in both ontological and theoretical capacities. We then trace a path towards a more robust justification for each type of simplicity principle, drawing on recent work in cognitive science and in philosophy of science, respectively. The resulting proposal hinges on the idea that simplicity is an evolved, virtuous cognitive bias—one that is a condition of our scientific understanding and, ultimately, of successful scientific practice. Finally, we make a case for why minimalists should take this proposal seriously, on the one hand; and for why generative linguistics would make for an interesting case study for philosophy of science, on the other.

✉ Giulia Terzian
giulia.terzian@gmail.com

María Inés Corbalán
inescorbalan@hotmail.com

¹ Department of Philosophy, University of Bristol, Bristol, UK

² School of Philosophy and Human Sciences, Federal University of Santa Catarina, Florianópolis, Brazil

³ Institute of Language Studies, University of Campinas, Campinas, Brazil

1 Introduction

Simplicity¹ is widely hailed across science and philosophy as a desirable trait of our theories, models, explanations, etc. Generative linguistics is no exception, on the contrary going so far as to elevate simplicity to the status of high priority research goal.² It is therefore striking, given the purported centrality of this notion, that generativists have yet to offer satisfactory answers to the fundamental questions of how simplicity is to be defined, measured, traded-off and—above all—justified. As we will argue, the latter worry in particular becomes even more pressing under the recent Minimalist Program (MP), which is predicated on the idea that simplicity is a fundamental and defining feature of the human language faculty, a key ingredient in linguistic explanation, and a prominent theoretical constraint (Chomsky 1995). In order to back up these claims, we begin by reviewing what we see as the most salient junctures in generative conceptualizations of simplicity, in Sect. 2.³ Among other things, this exercise will reveal that there continues to be a good deal of ambiguity concerning the alleged bearer(s) of this notion, thus explaining the widespread and well-documented (and otherwise unwarranted) expectation that the ontological and theoretical notions of simplicity should converge (Sect. 3). The second and main part of the paper is devoted to showing that the issues of justification and convergence become much more tractable as long as generativists embrace a more naturalistic methodology; importantly, our proposal will be conciliatory rather than antagonistic.⁴ We make this case in Sects. 4–6, by examining the notion of simplicity through the lens of a pair of recent debates in cognitive science and philosophy—respectively, on domain-general cognitive biases and on scientific understanding. Among other things we show that in its object-level capacity, simplicity is much more plausibly construed as a derived (or inherited) vs. intrinsic property of the language faculty. Moreover, we argue that minimalist appeals to simplicity as a theoretical value *can* be justified—as long as minimalists themselves

¹ And cognate notions such as parsimony, economy, elegance, naturalness, beauty, etc. Unless explicitly stated, any mention of simplicity should be interpreted as shorthand for ‘simplicity and cognate notions’. To be absolutely clear, we are *not* claiming that any or all of these notions are equivalent and interchangeable; we are merely using ‘simplicity’ as an umbrella term for the sake of discursive fluidity. Readers will find this important clarification borne out in the forthcoming discussion.

² Generative linguistics is typically construed as a branch of cognitive science, insofar as its chief concern is the study of the human language faculty, and the latter is a component of our cognitive system. The point of this extremely rough characterisation is to distinguish the study of language as an internal state of a biological organ from language as a socio-cultural object.

³ We wish to emphasise that this will not merely consist of a regurgitated version of agreed-upon facts: the chaotic, disconnected nature of the literature makes this near impossible. Sect. 2 is the result of our own laborious reconstruction of the history of simplicity in generative linguistics. We are not claiming that this is the best or only such reconstruction (for a critical analysis of generativism from the perspective of linguistic historiography, see Kertész 2010); nor that the stages identified therein are entirely clearcut or conceptually isolated from one another. Indeed, our discussion explicitly marks the continuities underwriting the evolving generative conceptions of simplicity.

⁴ We use the term ‘naturalism’ to refer to the methodological approach, or attitude, that explicitly encourages and values a frank dialogue between philosophy and the sciences, and furthermore acknowledges that such dialogue may require philosophers to defer to scientists’ expertise; see e.g. Nersessian (1987), Ankeny et al. (2011) and Soler et al. (2014). This is a salient clarification given that the term ‘naturalism’ is sometimes appropriated by generativists to refer to the so-called ‘Galilean stance’ (cf. Sect. 3).

adopt a more flexible perspective of the aims of scientific inquiry on the one hand, and of which epistemic vehicles can further such aims on the other. Section 7 concludes.

2 Simplicity in generative linguistics: a bird’s-eye view

2.1 Simplicity double-act: theory selection and grammar selection

Up until the mid-50s, the main goal of generative linguistics was to arrive at a descriptively adequate characterisation of human languages (Chomsky 1955, 1957). Simplifying greatly, this amounted to a two-fold task: formulating grammars—understood here as systems of rules—underlying existing languages, and producing a general *theory* of grammar. Accordingly, up until this point simplicity appeared in a purely methodological capacity, shaping the search for ‘best’ theory into the search for the theory of grammar that is ‘simplest’: more unified, containing fewer and shorter rules, and fewer symbols.⁵

The first salient juncture coincides with the explanatory turn of the 60s, as generativists direct their attention to the question of *how* individual linguistic agents learn, or acquire, (their native) language.⁶ Loosely put, the idea in these early stages of linguistic theory is that at birth, a speaker’s native language is underdetermined by the available evidence (external linguistic stimuli); language acquisition comes about as the speaker (or rather the speaker’s linguistic module) ‘chooses’ among possible grammars, eventually settling on the correct one. But how does our cognitive apparatus complete such a task, given the infinite size of this class? To obviate this difficulty,

For the construction of a reasonable acquisition model, it is necessary to reduce the class of attainable grammars compatible with given primary linguistic data to the point where selection among them can be made by a formal evaluation measure. (Chomsky 1965, p. 35)

Crucially, the task of ‘reducing the class of attainable grammars’ is now explicitly ascribed to the human language faculty. More specifically, on this early explanatory account it is postulated that humans are genetically endowed with a rich ‘universal grammar’—consisting of more or less abstract rules—which therefore curtails the space of ‘attainable grammars compatible with given primary linguistic data.’ This posited universal grammar thus turns language acquisition from an impossible to a feasible task; at the same time, it is not thought to achieve a definitive reduction of the space of possible grammars. That is, it is still thought that there can be more than one descriptively adequate grammar for a given language; and that given two descriptively adequate grammars, (part of) the role of the language faculty is to provide the procedure for selecting the ‘correct’ one. Chomsky then postulates that simplicity enters this very selection procedure; put differently, and only slightly more precisely, it is thought that

⁵ The origins of this grammar-specific simplicity criterion are found in (Chomsky 1951, p. 6): “the criteria of simplicity are as follows: that the shorter grammar is the simpler, and that among equally short grammars, the simplest is that in which the average length of derivation of sentences is least.”

⁶ At this stage, the terms ‘learn’ and ‘acquire’ were used fairly loosely and interchangeably.

some sort of simplicity metric (or ranking, or evaluation) is part of the actual process of language acquisition:

Here in outline is the device Chomsky used in the mid-1960s to make sense of how the child’s mind automatically ‘selects’ grammar X as opposed to Y—that is, learns X as opposed to Y, given data D. Think of X and Y as sets of rules, both candidates as descriptions of language L or, more carefully, of the data available to the child’s mind. Which [...] should the child’s mind choose? Introduce now an ‘internal’ simplicity measure: rule set X is better than Y to the extent that X has fewer rules than Y. (Chomsky 2009, p. 28).

Thus, simplicity makes its first ‘double’ appearance, in an object-level as well as a theoretical capacity. Moreover, the internal notion is thought to play a prominent role in language acquisition, roughly in analogy to the way that supra-empirical criteria intervene in underdetermination scenarios.⁷

2.2 Simplicity internalised: from internal metric to innate endowment

The next key turn comes about as the explanatory question is gradually sharpened into the formulation now known as Plato’s Problem: How do children acquire language given the poverty of data initially available to them?

A little more specifically, foremost on the research agenda at this stage is the challenge of explaining the following observed facts about linguistic acquisition and competence:⁸

- (P1) the homogeneity of language acquisition within and across linguistic communities;
- (P2) the relatively short time it takes children to acquire their native language, given the poverty of input data;
- (P3) the vast diversity of languages.

Ultimately, generative efforts to account for (P1)–(P3) crystallised into the so-called Principles & Parameters framework (Chomsky 1981). The P&P model paints the following picture of the human language faculty (FL).⁹ In its initial state (i.e., when we are born) FL is genetically equipped with two types of resources: a set of universal principles and a set of 2-valued parametrized principles. In this initial state—known as Universal Grammar (UG)—the parametrized principles are ‘switched off’; the classic analogy invoked in the literature is of a dormant switchboard.¹⁰ Prompted by linguistic stimuli from the environment, FL ‘sets’ the value of these parameters. Language

⁷ See also Sober (1975, Ch. 2) and Sober (1978).

⁸ Linguistic competence is sometimes described as a kind of knowledge (of the grammar of the speaker’s native language), although it remains an open question just what sort of knowledge might be at stake. For instance, Chomsky categorically and convincingly rules out that it be identified with propositional knowledge (knowledge-that), and suggests it is a kind of tacit knowledge.

⁹ The literature refers to P&P interchangeably as a model, a theory, an approach, a framework, or even a program.

¹⁰ UG is sometimes referred to as “the theory of the initial state of the language faculty” (Chomsky 1995, p. 12), or “the theory of the biological endowment of [FL]” (Chomsky 1995, p. vii); and sometimes simply

acquisition is what happens as more and more parameters are set, as a result of an optimal interaction between FL and the linguistic environment. Once all parameters have been fixed, the (idealized) native speaker has achieved linguistic competence, i.e. language acquisition is complete. Crucially, (P1)–(P3) receive an elegant and seemingly plausible explanation by the lights of this model.

Notice however the absence of any explicit reference to simplicity in the foregoing, either as a theoretical property or as an internal feature of FL; yet there is no doubt that generativists continue to entertain both assumptions. A plausible explanation is that P&P is thought to embody both constraints, thus foregoing the need to make either explicit. How so? We suggest the following interpretations. First, P&P is a simpler theoretical construction compared to its predecessor, in three respects:

- Ontological parsimony: a small number of abstract principles and 2-valued parameters replace a complex structure of specific rules;
- Unification: UG is universal in a stronger sense than its lower-case predecessor;
- Explanatory power: language acquisition is now a (comparatively) low-complexity task.

Secondly, FL itself instantiates three kinds of simplicity on the P&P account:

- Elegance and unification: the constituents of UG are fewer and highly abstract;
- Economy: FL operates more efficiently and with fewer resources.

2.3 From Plato's problem to Darwin's problem

The final turn coincides with the birth of the recent Minimalist Program (Chomsky 1993, 1995). MP takes as premises that the generative enterprise, up to and including P&P, has successfully addressed both the descriptive challenge (by identifying the particular grammars underpinning individual languages) and a first layer of the explanatory challenge (by producing a model of human language acquisition). As we've just seen, the gist of the latter is that FL develops or 'grows' from an initial, universal state—UG—to a steady state—the individual language/grammar—, prompted by environmental linguistic stimuli.

Minimalism explicitly seeks to address a second layer of the explanatory challenge, sometimes described as the challenge of arriving at a *principled* explanation of the properties of FL.¹¹ More specifically, informing the minimalist research agenda are the following questions:

- (M1) Exactly how does FL work?
 (M2) Why does it have the properties it has?

Footnote 10 continued

as "The biological equipment that makes language acquisition possible" (Boeckx 2010, p. 486). In this context the term 'theory' seems to be used rather loosely, then, to denote a cluster of constraints that pick out UG. See also Sect. 3.

¹¹ Several authors have noted that early hints of this idea can be traced all the way back to early generative writings; that is, the idea that generative linguistics should aspire to one day achieve this sort of explanatory depth far predates MP itself. For further discussion of the conceptual continuities spanning generative history, see e.g. Freidin and Lasnik (2011), as well as Boeckx (2006).

(M3) How could FL have evolved?

MP's key conjecture is that FL is a cognitive module that interacts with nearby modules (the sensori-motor and the conceptual-intensional systems) *in an optimal way*. This is the so-called Strong Minimalist Thesis:

(SMT) FL is an optimal solution to the interface conditions imposed by the conceptual-intensional and sensori-motor systems.¹²

Importantly, minimalist attempts to substantiate SMT rely heavily (and once again explicitly) on two notions of simplicity, one external and one internal.¹³ In fact, contrary to the official party line we find that extant discussions underwrite a more fine-grained taxonomy of simplicity:

- An external notion, labeled methodological economy (MS). This is the familiar—imprecisely defined—theoretical value, guiding linguistic inquiry (*qua* scientific inquiry).
- Two internal notions, typically lumped together under the label of ontological or substantive economy.
 - Procedural simplicity (PS): FL operations are subject to a number of economy constraints on derivations and on representations.¹⁴
 - Ontological simplicity (OS): UG is ontologically parsimonious, sparse, non-redundant.

This is MP's main gamble: that *FL is both procedurally and ontologically simple*. By way of investigating this conjecture, minimalist inquiry has largely focused on re-examining extant linguistic accounts by the lights of MS, PS and OS. To the extent that “Minimalist considerations motivate rethinking and replacing [previously accepted] assumptions and technical machinery” (Hornstein et al. 2005, p. xii) this can be seen as an attempt to address (M1). More recently, minimalists have turned their attention to (M2)—the demand for a principled explanation of FL-properties—and (M3)—known as Darwin's Problem (Boeckx 2009). We'll briefly expand on these in turn.

Once again simplifying greatly, we may see attempts to address (M2) as guided by the ‘third-factor hypothesis’: that at least some and perhaps most properties of FL may derive from, and be explained by “even more general, perhaps “language-external” principles” (Chomsky 2004, p. 24). This idea stems from Chomsky's suggestion that

¹² A more recent conjecture is that FL is an optimal solution only, or primarily, to C-I design specifications; see e.g. Chomsky (2007).

¹³ The two types or levels of simplicity underlying MP are also discussed by Boeckx (2006) in terms of elegance and beauty, respectively. Thus, he writes that “Work within minimalism seeks to develop beautiful theories” (2006, p. 120) where a ‘beautiful theory’, as we understand it, is one that offers a satisfactory principled explanation of linguistic phenomena, as required by (M2) above.

¹⁴ Economy conditions on derivations and representations guarantee that the latter are optimal or (computationally) efficient in some—more or less precisely specified—sense. Various such conditions have been proposed in the minimalist literature, including e.g.: Inclusiveness; Shortest Move; Last Resort; Procrastinate; Greed; Enlightened self-interest; No-tampering; Full Interpretation. Details don't really matter here; the general idea is that grammars are organized frugally to maximise resources. For a (very) critical discussion of economy principles proposed under MP, see Lappin et al. (2000).

the growth of language in the individual is determined by the interaction of three factors: (a) genetic endowment; (b) experience; and (c) general principles not specific to the language faculty. (Al-Mutairi 2014, p. 73)

What might these principles be? Beyond the fact that they are non-domain-specific, universal, and language-external, opinions on this matter diverge.¹⁵ We are more interested in the fact that the hypothesis marks a fundamental shift in the allocation of explanatory burden. Recall the cardinal hypothesis of P&P: that UG—our genetically determined linguistic endowment—is rich enough to bear the explanatory bulk of language acquisition (and the workings of FL, more generally). By contrast, under MP it is thought that

a “principled explanation” of the language faculty and its properties may be achieved by “shifting the burden of explanation from the first factor [...] to the third factor” (Chomsky 2005, p. 9). (Al-Mutairi 2014, p. 75)

Crucially, the rationale for such a shift comes from the generative community’s more recent concern over reconciling models of FL with evolutionary theory. For, while P&P offers an attractive answer to Plato’s Problem as a result of countenancing a rich, genetically encoded UG, this very assumption makes it problematic from an evolutionary perspective—particularly given the relatively short time that language has ‘been around’ (less than 100,000 years by most estimates).¹⁶ This is Darwin’s Problem. In response, minimalists have adopted a two-pronged simplicity-based strategy, devised to ease the evolutionary pressure on FL and thus avoid having to posit ‘multiple miracles’: on the one hand, shift the burden of explanation from the first to the third factor; on the other, seek to ‘empty’ UG as much as possible, either by eliminating entities outright or by reducing them to a thinner and more fundamental ontological basis.

The picture that emerges from the foregoing (lamentably brief) overview could be described at once as dynamical and volatile. We’ve seen the notion of simplicity occupy a central role throughout the history of generative inquiry, albeit under rather changeable guises. In particular, we saw it double up as theory-level and object-level property fairly early on, before mutating further still—most recently, into what we have labelled PS and OS—in this latter capacity. What we have not seen—what is

¹⁵ The term ‘third factor’ is introduced by Chomsky (2005) (although the germ of the idea may have already been present in 1965), and characterised as consisting of “Principles not specific to the faculty of language,” including “(a) principles of data analysis that might be used in language acquisition and other domains; (b) principles of structural architecture and developmental constraints [...] including principles of efficient computation” (Chomsky 2005, p. 6). Elsewhere, Chomsky occasionally seems to think that principles of type (b) might include general laws of physics. We’ll disregard the latter interpretation: partly because it is comparatively underdeveloped, and partly because the former is more attuned with the focus of our discussion.

¹⁶ In this connection, Boeckx draws attention to certain similarities between the acquisition and evolution challenges. Both raise a problem of reconciling a complex phenomenon—respectively, language acquisition at the individual level and evolution of FL at the species level—with the strict temporal constraints to which the phenomenon is subject. This observation in turn motivates adopting analogous solution strategies, Boeckx suggests: “the way we should try to address and solve [the evolution challenge] is to do exactly what we have done for Plato’s Problem, namely to [...] make sure that the thing that has to evolve is actually fairly simple” (2009, p. 47). We merely add that whether the proposed approach—namely, to aim for parallel solutions based on certain similarities between the respective problems—will work is ultimately an (as yet open) empirical matter, and certainly not one we are equipped to pronounce ourselves on.

remarkably absent from the literature and not just our overview—is a corresponding, parallel narrative as to *why* we should take these simplicity ascriptions at face value. This is true not just of theory-level simplicity claims, for which robust justifications are notoriously hard to pin down in general. It is also and much more pointedly true of their object-level counterparts. As noted at the outset of the paper, this is a puzzling situation given the centrality of the idea that simplicity is a property of FL, both throughout generative history and most explicitly under MP. Indeed, in light of this latter fact the lack of a solid justificatory basis for either kind of simplicity claim becomes a legitimate and serious concern. Happily, we think there is a way to mitigate both worries, as we'll see in Sects. 4–6. Before we do so, the next section briefly expands on an additional important confounding factor in generative discussions of simplicity, witnessing a sustained conflation between theory- and object-level notions on the one hand, and an expectation that the two should converge on the other.

3 Galileo meets Ockham: the purported convergence of simplicities

Patently, simplicity concerns have been a constant fixture in the development of generative linguistics. By contrast, the interpretation of this notion has fluctuated considerably from one framework to the next, and sometimes within one and the same framework. More worryingly, discussions of simplicity are mired in at least one important sort of ambiguity, between ascriptions of simplicity to the object of study and to linguistic theory itself. A representative example of this sort of confusion is found in the following passage:

To repeat, minimalism is a project: to see just how well designed the faculty of language is, given what we know about it. It's quite conceivable that it has design flaws, a conclusion we might come to by realizing that the best accounts contain a certain unavoidable redundancy or inelegance. (Hornstein et al. 2005, p. 14)

In fact, the conflation of theory- and object-simplicity is but one instance of a more general trend, within the generative community, of failing to disambiguate between theory and object *simpliciter*. Particularly notable instances of this tendency are the notions of 'grammar' (cf. 2.2) and, later on, UG. Thus, for instance, UG is described simultaneously as an object of linguistic inquiry—specifically, the system of universal constraints that constitute our innate linguistic endowment—and as the theory of that same object—i.e. the theory of the initial state of FL. This poses a non-trivial interpretation problem, for instance when it comes to understanding the linguist's directive to 'rethink the structure of UG', or 'minimise UG'.¹⁷

Acknowledging this conflatory habit affords us an intuitive grip on the minimalist expectation that theory- and object- simplicity should converge. We suggest that this convergence assumption can be further unpacked in terms of the following explanatory factors: (E1) a largely implicit commitment to a strong form of (semantic scientific)

¹⁷ Similarly, in the discipline's early days the term '(generative) grammar' was used ambiguously to refer both to the linguist's object of study (i.e. a grammar for a particular language), and the linguist's theory (i.e. the theory of generative grammar).

realism, (E2) a commitment to a metaphysical thesis according to which the world is simple (known in generative circles as the Galilean stance, or style), (E3) a commitment to a ‘naturalist’ stance according to which ‘language should be studied in the same way as any other aspects of the natural world’ (Al-Mutairi 2014, p. 34), (E4) a commitment to the ‘Occamist urge to explain with only the lowest number of assumptions’ (Boeckx 2010, p. 494), (E5) a failure to clearly distinguish between (E1)–(E4).

Illustrations of (E1)–(E5) are anything but difficult to find in the literature. Here are just a few representative passages:

We construct explanatory theories as best we can, taking as real whatever is postulated in the best theories we can devise (because there is no other relevant notion of ‘real’), seeking unification with studies of other aspects of the world. (Chomsky 1996, p. 35) (as cited in (Smith and Allott 2016, p. 204))

[What] further properties of language would SMT suggest? One is a case of Occam’s razor: linguistic levels should not be multiplied beyond necessity, taking this now to be a principle of nature, not methodology, much as Galileo insisted and a driving theme in the natural sciences ever since. (Chomsky 2007, p. 16)

[The] Galilean style [...] is the central aspect of the methodology of generative grammar. [...] The Galilean program is thus guided by the ontological principle that “nature is perfect and simple, and creates nothing in vain” [...]. This outlook is exactly the one taken by minimalist linguists. [...] The road to Galilean science is to study the simplest system possible [...]. (Boeckx 2010, p. 498)

Without adhering to the Galilean style, without the strongest possible emphasis on simplicity in language (the strongest minimalist thesis), it is hard to imagine how we might ever make sense of the properties of FL. (Boeckx 2010, p. 501)

Notice the no-miracle flavour of the last quote; paraphrased from context, it amounts to the following: If FL weren’t as MP describes it, (i) the success of MP would be a miracle and (ii) the evolution of FL would require multiple miracles. Interestingly, this parallels the argumentative strategy employed in justifications of a rich, innate UG (cf. also footnote 16). Paraphrasing from Al-Mutairi (2014) (and his paraphrase of Chomsky): Factor I must be non-empty (‘*something* must be special to language’) or else language acquisition would be a miracle; Factor III must be non-empty or else language evolution would be a miracle.

The foregoing sections have sought to unearth the many faces of simplicity in generative linguistics. Perhaps the most salient aspect of the resulting picture is a persistent and *indiscriminate* pull towards simplicity—an entrenched belief that simplicity colours both theory and object of study—that sits on a shaky foundation, captured by (E1)–(E5) above. In light of these facts, it is therefore hardly surprising that justification questions have been largely overlooked. In the next sections, we offer the minimalist a way out.¹⁸

¹⁸ One of the referees remarked that, 25 years down the line, the minimalist community’s interest in the original research program (i.e. arriving at an understanding of FL that goes beyond ‘mere’ explanatory adequacy, which would require *inter alia* seriously investigating the third factor hypothesis) has

4 Taking the third-factor hypothesis to the next level

We see the rise in prominence of the third-factor hypothesis as one of the most promising aspects of recent minimalist inquiry. At the same time, it is our impression that its significance and potential ramifications have thus far been under-appreciated by the minimalist community.¹⁹ In large part, this is because generative linguistics has not quite lived up to its own self-identification as a branch of cognitive science, at least insofar as it has foregone substantive engagement with said discipline. In this section we make a case for the importance of a collaborative dialogue between linguistics and cognitive science: not just for the sake of honouring the former’s naturalistic commitment (although this would be a good enough reason by itself); but also, more pointedly, as a way to address and mitigate the justification worry with respect to object-simplicity claims. In light of a cluster of well-supported findings in cognitive science, we’ll see, the long-standing generativist ‘hunch’ that FL is in *some* sense simple stands a good chance of being vindicated.

To see how, recall first that minimalists have sought to substantiate SMT by placing a premium on OS as a guide to constructing models of FL (Sect. 2.3). Such models thus witness a reduction of both the innate and the domain-specific content previously assumed to be part of UG. Moreover, while the implementation of OS sometimes results in the outright elimination of entities from UG, more often it leads to a relocation of content, either from UG to other cognitive systems (third factor), or from UG to the environment (second factor), or both. Crucially,

1. at least part of the content relocated to other cognitive systems consists of PS constraints;
2. to the extent that SMT is true, content that is relocated to other cognitive systems is still ‘part of’—or accessible to—FL.²⁰

Footnote 18 continued

gradually dwindled, as minimalists have returned to more parochial (descriptive and explanatory) work on particular languages (or: grammars). The referee thus wondered whether there is still a minimalist community ‘out there’ that could fill the role of our audience. Put differently, who is ‘the minimalist’ we are purportedly addressing? Our response to this interesting observation is threefold. On the one hand, if the community has indeed drifted away from the program’s original research goals, then—seeing as several of these goals have yet to be met—that is all the more reason to put these suggestions to them. We don’t for a moment presume that our paper could be so impactful as to single-handedly turn the minimalist tide, of course. Still—and here is our second point—there are at the very least scattered individual researchers out there who remain invested in the original questions. However few in number, they are a worthy audience for this and like-minded papers. Finally, as we write in Sect. 6 we are in fact addressing more than just ‘the minimalist’. To borrow Kitcher’s (2019) terminology, we hope to address the Scientists (cognitive scientists as well as generativists), the Philosophers (especially those interested in the role of aesthetic values in scientific contexts), and last but not least the Interested Citizens (or more prosaically, any and all those who recognise the value of the sort of interdisciplinary investigation we are promoting).

¹⁹ With (isolated) exceptions, of course. As we’ll see below, some minimalists (e.g. Boeckx 2016) have embraced the methodological shift prompted by the third-factor hypothesis more explicitly than others.

²⁰ The widely cited paper by Hauser et al. introduces a distinction between the language faculty in a broad sense (FLB)—which includes the interface systems—and in a narrow sense (FLN)—comprising just the recursive computational system. The authors’ hypothesis, which enjoyed a rapid uptake within the generativist community, is that “FLN [...] is the only uniquely human component of the faculty of language” (2002, p. 1569), whereas FLB is (plausibly) shared with other species. In other words, on this hypothesis: *some* language-specific content exists; and all (and only) such content is confined to FLN.

More plainly: taken together, SMT and the third-factor hypothesis entail (among other things) that simplicity is no longer a domain-*specific* property of UG, but rather a *domain-general cognitive feature*. Oddly, minimalists have largely downplayed or even ignored the ramifications of this fact, nor have they ventured to seek its corroboration (or correction) from empirical evidence.²¹

We think such evidence can be found in recent empirical studies conducted by cognitive scientists of different ilk, united by the project of investigating simplicity as a general principle of cognition. The central hypothesis driving these studies is that *our cognitive system favours simple interpretations* (mental models/hypotheses) *of the data*; put differently, we are wired to search for simple patterns in the world. We'll refer to this as the cognitive simplicity hypothesis (CSH).

What makes a pattern, or a hypothesis, simple? Typically, cognitive scientists employ an information-theoretic measure of simplicity (e.g. as provided by Kolmogorov complexity theory, or Shannon's information theory) in a universal coding language. The general idea is that the simplicity of a pattern can be measured by the extent to which it compresses—provides a compact encoding of—the data; the simplest pattern, corresponding to the shortest coding, provides the least redundant representation of the data.²²

Thus far, CSH has been vindicated by a host of empirical studies from various sub-domains²³ showing that this increasingly well-documented simplicity bias supports

²¹ Once again, we do not claim that *no* minimalist has taken the third factor hypothesis and its consequences seriously; only that this is true of the community, at a programmatic level. For instance, one of the reviewers pointed out that Boeckx (2014b, 2016) clearly favours the idea that third factor content—under the heading of FLB—could play a non-trivial role in language acquisition, and indeed engages with literature from cognitive science to discuss specific (possible) illustrations of such content. Indeed, in more than one place Boeckx is explicit about his commitment to the 'biolinguistic enterprise', understood as "The road leading theoretical linguistics beyond explanatory adequacy, that is, towards a naturalistic, biologically grounded, better-integrated cognitive science of the language faculty" (2014a, p. 1). In this respect, we are very much on the same page as Boeckx: we too are pushing for genuinely interdisciplinary collaboration between the linguistic and cognitive science communities; we too regard the third factor hypothesis as very much worth investigating from an integrated cognitive perspective. What sets apart our proposal from his hinges on our respective foci. Put briefly and certainly not exhaustively, in his more recent work Boeckx seeks to explore the consequences of the third factor hypothesis on what we might term a more general scale: as a means to (re-) address Plato's Problem by effecting a 'fairer' redistribution of the explanatory burden between FLN (comprising perhaps no more than a universal, invariant computational system) and FLB (comprising now also domain-general learning mechanisms). Our focus in this subsection is in a sense much narrower, and informed by somewhat different premises (something not altogether surprising since, as philosophers, we are 'outsiders' to the linguistic debate). We are interested specifically in the 'third factor consequences' for the (internal) notion of simplicity; in a sense, we want to see 'what happens' to this central notion if and once an interdisciplinary approach is implemented. To our knowledge, *this* particular slant remains under-discussed in the generativist literature (though not within the cognitive sciences, as we'll see). That said, there is also a sense in which the approach we are promoting is broader in scope than Boeckx's. Like him, we pressing for a serious collaboration between linguistics and cognitive science, *with respect to their shared subject matter*. But we are advocating just as strongly for a collaboration with philosophy of science, *with respect to the aims, methods and vehicles of linguistic inquiry* (see Sects. 5–6).

²² The invariance theorem (Li and Vitányi 1997) ensures that the shortest description of any object is language-invariant (up to a constant).

²³ Including: concept learning (Feldman 2003); perceptual organization and category learning (Gershman and Niv 2013; Pothos and Chater 2005); function acquisition (Narain et al. 2014); causal reasoning (Lombrozo 2016; Bonawitz and Lombrozo 2012); sensorimotor learning (Genewein and Braun 2014). For a recent survey, see (Feldman 2016).

successful explanations and predictions. From this vast literature, we single out for mention a handful of studies that focus on the role of simplicity in language learning/acquisition (Onnis et al. 2002; Hsu et al. 2013; Chater et al. 2015) and language evolution (Christiansen et al. 2006; Chater and Christiansen 2010; Culbertson and Kirby 2016), and present what we regard as their key highlights and points of contact with minimalist inquiry.

Recall the generative solution to the acquisition problem: a language faculty endowed with a rich, innate UG. This has two crucial explanatory benefits: it accounts for the universality of language, and it ‘compensates’ the paucity of data available to the child. The latter is a central ingredient of the so-called ‘poverty of stimulus’ argument for UG, which emphasises that said data is not only quantitatively limited, but also almost entirely positive, thus making the putative task of learning language from data alone implausibly hard, if not impossible. While the argument—which is cast as an instance of inference to the best explanation—continues to hold sway among generativists, recent empirical studies on language learning point to a way out of the problem of positive evidence. In a nutshell, one of their key conclusions is that in the presence of a general cognitive simplicity principle, the input data *is* sufficiently rich to ground language acquisition. The significance of this result cannot be understated, we think: if CSH continues to hold up under future empirical scrutiny, it would seem that the acquisition problem could be put to rest *without needing to postulate any innate linguistic content*.

Indeed, if the above results indicate that we *can* do without innate linguistic content, a second set of studies suggest that we *should* forego such assumptions. To see this, recall the minimalist strategy to address Darwin’s Problem: shift content from the first to the third factor, and empty UG of any redundant content. While this is promising from a naturalistic perspective, at least insofar as it is intended to align linguistic theory with evolutionary theory, we suggest that, in light of the following, minimalists as a community can and should take their strategy one step further.

Suppose we ask: what’s left in FL once any and all redundant content is stripped away from UG? Minimalist answers will vary (even significantly), but most will make reference to at least one specific *linguistic* property, or mechanism; in the terminology borrowed from the cognitive sciences, a domain-specific hard constraint. In the best-case minimalist scenario, only one such constraint would be required to explain—in conjunction with more general cognitive mechanisms—everything from language acquisition to language evolution. However, even this ideal model proves problematic from an evolutionary standpoint. The problem is that a domain-specific hard constraint, of the sort that would qualify as first-factor content, is unlikely to have evolved—even more so given the relatively recent appearance of language. On the other hand, the evolution of domain-*general*, weak constraints (or biases) seems well-supported by evolutionary theory. In particular, there seems to be mounting evidence to the effect that one such constraint is none other than the cognitive simplicity principle.

Against this backdrop, a number of recent studies have set out to investigate the conjecture that language may be the result, not of a specific evolutionary adaptation²⁴

²⁴ Nor of a ‘miracle’ or an evolutionary jump, as some minimalists have occasionally suggested. For recent discussion of a saltationist account of language evolution, see Kinsella (2009), Di Sciullo and Boeckx (2011) and Tallerman and Gibson (2012).

but rather of the interplay of evolved, weak biases and cultural evolution. One such argument is made by Culbertson and Kirby (2016), who start off by distinguishing two ways in which a property may be specific of a given cognitive domain: the property may have evolved for a specific functional purpose, or it may have evolved for either a different or a domain-general purpose, eventually coming to interact with a specific cognitive system in a unique way. The authors then argue that language evolution²⁵ is most plausibly captured by the latter explanatory route. Their argument draws on two main sets of results, obtained via computational models of language evolution. The first set shows that a genetically determined universal grammar—the sort of innate content posited by generative theories—is unlikely to have evolved, either by natural selection or by other evolutionary mechanisms.²⁶ The second set suggests, first, that cultural evolution has an amplifying effect on weak cognitive biases; secondly, that “weak biases for language learning are *more* evolvable by virtue of cultural evolution’s amplifying effect” (2016, p. 3). From the foregoing, the authors correctly draw the cautious conclusion that,

While this does not categorically rule out the existence of very strong (or inviolable) biases that have evolved specifically for language, it clearly suggests we should not treat them as the default hypothesis. (2016, p. 9)

More interestingly still, they make a compelling case for the hypothesis that several linguistic phenomena could be domain-specific *effects* (vs. hardwired constraints) of a domain-general simplicity bias as the latter interacts with ‘linguistic representations’—that is, in the terminology of the previous sections, with second-factor content, i.e. the linguistic environment (see also Thompson et al. 2016).

Where does the foregoing leave us? Earlier we noted how recent attempts to flesh out SMT have led minimalists to place more weight on the third-factor hypothesis. However, ensuing proposals have struggled to genuinely distance themselves from the dominant model of FL as a language-specific module structured by innate, language-specific hard constraints. While this is certainly understandable from a sociological perspective, it seems unsatisfactory by naturalistic standards. This becomes starkly evident once we take into account the vast array of empirical studies that point, rather convincingly, to the implausibility of said model; and which furthermore offer an alternative, scientifically robust framework within which solutions to both Plato’s and Darwin’s Problems appear well within reach.

In the next two sections we push for an analogous ‘naturalizing’ move with respect to theory-level simplicity claims. One of its main upshots will also mirror an important takeaway from the present section: namely, that serious pursuit of a justification of theory-simplicity may require breaking down inter-disciplinary barriers and—in this specific instance—looking at what philosophers have to say.

²⁵ That is, the evolution of “the linguistic system—its architecture, the representations it operates on, the constraints it is subject to” (Culbertson and Kirby 2016, p. 1).

²⁶ Such as the Baldwin Effect, “whereby traits that were previously acquired through experience become nativised” (Culbertson and Kirby 2016, p. 2).

5 Theory-simplicity: a compatibilist alternative

As noted in earlier sections, and bracketing issues of object-theory conflation for the moment, generative appeals to simplicity as a theoretical virtue have sought to fall in line with a general tendency, in science and philosophy, to favour simple theories, models, explanations, etc. However, such appeals have rarely been accompanied by in-depth reflection on the questions of how theory-simplicity ought to be defined, measured, justified, and traded-off.²⁷ In fairness, generativists are hardly the exception in this regard; nonetheless, given the prominence ascribed to simplicity in minimalist theorising, we suggest such a reflection should be delayed no further.

To this end, a natural source of inspiration is the philosophical discussion on the role of theoretical values in scientific practice. Within this debate, analyses of so-called aesthetic values—including simplicity—traditionally fall into one of two camps: those that construe aesthetic values as ‘merely’ *pragmatic* criteria, and those that ascribe a more substantive, *epistemic* role to these notions.²⁸ Construals of the first sort typically place a strong emphasis on the variability, relativity and even subjectivity of aesthetic (and any other non-evidential) values; on this view, simplicity is cast in a strongly instrumentalist light, with connotations of ‘easy to use’, and the like. By contrast, accounts of the second sort regard all such values as truth-conducive—albeit to different degrees, with greater weight being allocated to evidential criteria such as empirical adequacy and predictive power.²⁹

Here we sketch a compatibilist alternative to the above, that draws on recent proposals according to which aesthetic values do indeed serve a substantive epistemic function in scientific practice, without however relinquishing their pragmatic connotation (Breitenbach 2013; de Regt 2017; Kosso 2002; Ivanova 2017). More specifically, on this view aesthetic values are epistemically ‘active’ insofar as they are indicative of, and conducive to, *understanding* (of relevant target phenomena), where the latter is a central aim of science. Our main contention is that an analogous recalibration of the aims of inquiry would be both recommendable and potentially fruitful in the generative context. Given that this move hinges in turn on the epistemic notion of understanding, some stage-setting is appropriate at this point. We give a brief overview of the ongoing philosophical debate surrounding the notion of understanding as an aim of science in Sect. 5.1, which we then tie in with discussions of simplicity in 5.2. Against the resulting backdrop, we then comment on two extant analyses of simplicity in the context of generative linguistics (Sect. 5.3). Ultimately, we’ll see that neither is entirely satisfying precisely because they fail to distance themselves from the traditional adversarial narratives of science as either a truth-bound enterprise, or as subject to mere empirical adequacy standards. Building on the foregoing, Sect. 6 outlines what the proposed methodological and philosophical shift might ‘look like’ in generative linguistics.

²⁷ In what follows, unless otherwise indicated any mention of simplicity should be understood to refer to theory-simplicity.

²⁸ See e.g. Baker (2003), Barnes (2000), Schindler (2018) and Van Fraassen (1980).

²⁹ To repeat, generativists have very rarely engaged with this debate; to our knowledge, the only two exceptions are Barrios (2016) and Ludlow (2011), more about which in Sect. 5.3.

5.1 Scientific understanding: a brief overview of the debate

Understanding is currently (and has been for the past decade or two) a hot topic both within general epistemology and in the philosophy of science. While the landscape of this philosophical debate is heterogenous with respect to what we might label ‘local’ issues (more about which shortly), it is fair to say that there is a broad consensus according to which understanding is a cognitive-epistemic achievement which is (i) *more demanding than knowledge*; and (ii) tightly enmeshed (if not identical) with the central scientific aim of *producing explanations* of natural and social phenomena.

Both (i) and (ii) are fairly nebulous as they stand, of course. Unsurprisingly, disagreements have arisen wherever attempts to sharpen either thesis have been made. This is true more pointedly of (i): here, the issue that has proven to be particularly divisive concerns the relation of understanding to knowledge. More specifically, the main sticking point is whether understanding is a subspecies of knowledge, or if the two are entirely distinct epistemic achievements (see e.g. Grimm et al. (2017) for an excellent introduction). Within the former camp, moreover, further disagreement concerns whether understanding and knowledge share some vs. all of the same satisfaction conditions (minimally: truth, belief, justification).³⁰

In what sense, then, is there any kind of agreement over (i)? The consensus is that understanding requires ‘something extra’ over and above knowledge: namely, it requires that the subject *grasp* (at least some among) the salient explanatory connections within the domain that is the target of understanding. For instance, Kvanvig writes that

One can know many unrelated pieces of information, but understanding is achieved only when informational items are pieced together by the subject [...]. [This is the] crucial difference between knowledge and understanding: that understanding requires, and knowledge does not, an internal grasping or appreciation of how the various elements in a body of information are related to each other in terms of explanatory, logical, probabilistic, and other kinds of relations [...]. (2003, p. 192f.)

More generally, different authors have offered slightly different characterisations of the notion of grasping.³¹ By and large however it is agreed that grasping is not reducible to propositional attitudes such as knowledge or belief. We follow Bailier-Jones (1997) and Reutlinger et al. (2017) (who in turn seem to express an implicit consensus in the literature) in allowing that grasping is philosophically primitive, though not scientifically so. Thus, insofar as it is a cognitive activity, grasping is a legitimate object of study for the cognitive sciences; but philosophically, it seems perfectly acceptable to

³⁰ For instance, Grimm and Khalifa take understanding to be factive, citing different arguments for this claim (Grimm 2006; Khalifa 2013). By contrast Elgin and Zagzebski side with the non-factivist camp (Elgin 2017; Zagzebski 2001). Similarly, on whether understanding requires at least one among belief, justification, anti-luck conditions see for example Grimm (2006), Kvanvig (2009) and Dellsén (2017).

³¹ For instance, some push for a conception of grasping as a sort of ability, e.g. the ability to manipulate the relations between propositions. See among others Hills (2016).

have the buck stop here.³² Importantly, grasping is acknowledged to be independent of truth, even as the (non-) factivity of understanding continues to be hotly debated. This issue is more salient than others, in the context of our discussion, because it speaks to the question of whether only true (or probably or approximately true) scientific theories are to be considered reliable vehicles of scientific explanation and therefore understanding, or whether other kinds of vehicles might be included in this class. This takes us back to item (ii).

The central questions here are, first, what counts as an explanation—of the sort produced by scientists in their effort to advance their (individual and/or collective) understanding of the world. The second question is whether understanding can be mediated by different epistemic vehicles (beyond theories in the traditional, propositional sense) or is instead restricted to a specific subclass of such vehicles. On both these counts, the literature offers a picture that is more distinctively pluralistic than divisive. Thus, more or less peacefully co-existing in the current landscape are those who argue that understanding can be yielded by causal, how-actually explanations (Khalifa 2017); non-causal, how-actually explanations (Lipton 2009); how-possibly explanations (Reutlinger et al. 2017); successful classifications (Gijssbers 2013); non-propositional representations (de Regt 2017); models and idealizations (Elgin 2007; Strevens 2016); and perhaps fictions and more besides (Lawler 2019).

5.2 Theoretical values and scientific understanding

What does simplicity have to do with the foregoing? The consensus view that emerges from the literature is that aesthetic values, alongside more ‘canonical’ values such as consistency or predictive power, play an often crucial role in the subject’s achievement of understanding of the target via one or more relevant epistemic vehicles. Crucially, they contribute to this epistemic goal precisely in virtue of their pragmatic dimension.

One way to flesh out this idea is via de Regt’s notion of intelligibility of scientific theory. In line with the above-mentioned literature, de Regt (2009, 2017) identifies as a central aim of science what he calls ‘understanding a phenomenon’, or UP: the understanding that is provided by having an adequate explanation of the phenomena being investigated.³³

Understanding (i.e. UP) is thus a relation between subject and world; crucially for de Regt, it is mediated by *intelligible theories*, where intelligibility is defined as

the value that scientists attribute to the cluster of qualities of a theory (in one or more of its representations) that facilitate the use of the theory.³⁴ (2017, p. 40)

³² Only one attempt has been made to further analyse the concept of grasping, that we know of, by Janvid (2018).

³³ The notion of explanation that de Regt has in mind is more flexible than traditional—for instance, strictly causal—conceptions. In other words, de Regt is among those who subscribe to a pluralist view of explanation (and of epistemic vehicles); in particular, on his view “all explanations are, in a broad sense, arguments. An explanation is an attempt [...] to provide understanding of the phenomenon or the situation by presenting a systematic line of reasoning that connects it with other accepted items of knowledge” (de Regt 2017, p. 25). We are sympathetic to de Regt’s conception, above all because we are sympathetic to its pluralist spirit.

³⁴ Two things should be noted here. The first is that de Regt’s discussion concerns the broader class of theoretical values, including but not limited to aesthetic values; for instance, he notes that “Causal structure

Notice that while UP has a distinctively epistemic ring to it, intelligibility has expressly pragmatic overtones. De Regt's key thesis is that the latter is a necessary condition for the former: that is, successful explanations of phenomena require intelligible theories. Therefore, since theoretical values help shape intelligible theories, they are themselves preconditions of explanatory understanding.

By explicitly recognising that the epistemic and the pragmatic dimensions are thus enmeshed, the perspective developed by de Regt and others is a dynamical one, certainly compared to more established, incompatibilist construals. Indeed, a distinctive and shared feature of the former is the importance ascribed to context in shaping UP, by acknowledging the variability of theoretical values and their respective weights along multiple dimensions: through history, across domains of inquiry, between scientific communities; and among members of these communities, depending on “background knowledge, metaphysical commitments, and the virtues of already entrenched theories” (de Regt 2009, p. 31).³⁵ Crucially, this multifaceted context-sensitivity doesn't collapse into relativism: as Douglas (2013, p. 802) puts it, “the proof will be in the pudding [...], and the pudding is relatively straightforward to assess. [...] We will be able to tell readily if the instantiation of a pragmatic-based value in fact proves its worth.”

One of the many merits of de Regt's account is that it pays the history of science its due attention, offering detailed case studies (mainly from the history of physics) as a means both to illustrate his proposal, and to ensure it remains tethered to scientific practice.³⁶ However, while de Regt makes a compelling case for a robustly contextualist account of theoretical values, we find that he ends up obscuring a particularly interesting fact as a result: namely, that while many theoretical values have come and gone over the course of the history of science (e.g. visualizability), the cluster of so-called aesthetic values has remained a more or less stable fixture throughout. This observation is one of the premises of Breitenbach's account, to which we now turn.

Like de Regt, Breitenbach argues that understanding is a ternary relation between theory, world and scientist; more specifically—with an emphasis that sets her apart from de Regt—the *scientist's cognitive structure and capacities*. Following the declared Kantian inspiration of her account, Breitenbach construes aesthetic judgments in science as second-order responses to “our awareness of the suitability of our intellectual capacities for making sense of the world around us” (2013, p. 92). Impor-

Footnote 34 continued

is a quality that is often regarded as enhancing the intelligibility of theories” (de Regt 2017, p. 109). Secondly, as with the notion of explanation, de Regt favours an interpretation of ‘theory’ that is loosened to encompass also models, idealizations, experimentations, etc.

³⁵ Forster and Sober (1994) also argue for a local justification of simplicity, from rather different premises. Their argument, very much boiled down, runs as follows: the main goal of model selection is predictive accuracy (rather than probable truth); insofar as simplicity minimises the risk of overfitting the data, it also favours predictive accuracy; therefore, simplicity should be favoured—in the context of model selection problems. See also (Sober 2002).

³⁶ Three of the final chapters of (de Regt 2017) are devoted to the discussion of, respectively: the intelligibility of Newton's gravitation theory; the role of mechanical models as vehicles of understanding in 19th century physics; the role of visualization—as a criterion of intelligibility and therefore a condition of explanatory understanding—in the transition from classical to quantum physics (in particular, its role in the Heisenberg/Schrödinger debate over the superiority of matrix mechanics versus wave mechanics, respectively, as means to understand atomic phenomena).

tantly, aesthetic judgments are thus neither directly about the world, nor about the theory *per se*. Rather, they are “essentially self-reflective,” in that they reveal—mark our awareness of—the attainment of a certain harmony between our cognitive makeup and the world, mediated by our representations (theories, models) of the latter. Therefore, *aesthetic values are conditions of understanding*. Moreover, insofar as this is the case we are also *justified in pursuing simplicity*, unity, beauty etc. in our theories: for, while it is neither necessarily nor contingently true that simple theories will provide understanding (much less be truth-conducive), nonetheless they

condition the possibility of such understanding, [and] providing such understanding is an essential requirement for any successful theory. (Breitenbach 2013, p. 96)

Together, Breitenbach’s and de Regt’s proposals offer a powerful and compelling account of the role of aesthetic values, including simplicity, in shaping scientific practice. Moreover, as we’ll see in Sect. 6, the conception of scientific practice (specifically, its aims and methods) underlying these and similar accounts offers a novel and fruitful vantage point from which to re-examine linguistic practice.

5.3 Barrios and Ludlow on simplicity in generative linguistics

To complete our stage-setting operation we now examine two separate discussions of theory-simplicity in the philosophy of linguistics offered by, respectively, Barrios (2016) and Ludlow (2011) (see footnote 29). In so doing we hope to further elucidate the merits of our preferred, alternative construal of this notion. The first thing to note is that both Barrios’s and Ludlow’s analyses are to a certain extent entirely compatible with, in particular, de Regt’s account of simplicity (among other aesthetic values). In particular, both authors agree that ascriptions of theory-simplicity are sensitive to contextual factors, in the sense that they vary from one scientific community to another, between stages of inquiry and scientific periods, and over time.³⁷

However, whereas Barrios correctly recognises and indeed emphasises the varied epistemic roles played by simplicity considerations *vis-à-vis* the explanatory aims of science, Ludlow strongly downplays (indeed, ignores) the connection between the pragmatic character of simplicity and the epistemic function it serves in contexts of theory construction, choice etc. Thus, Ludlow argues that simplicity, as this notion applies to scientific theories (as opposed to subject matter) in general, and linguistic theories in particular, is nothing more than a pragmatic criterion, narrowly construed as synonymous with ‘easy to use’: “when we look at other sciences, in nearly every case, the best theory is arguably not the one that reduces the number of components from four to three, but rather the theory that allows for the simplest calculations and greatest ease of use” (Ludlow 2011, p. 158).³⁸

³⁷ Cf. for instance Ludlow’s Theses I–III (2011, pp. 161–162).

³⁸ Ludlow very briefly acknowledges that alongside theory-simplicity, MP is also motivated by a second notion whose role is essentially that of an explanatory goal: namely, to reduce the subject matter of linguistics to one that is more fundamental (“low level biophysical processes” (Ludlow 2011, p. 160); but cf. footnote 15). This is of course the interpretation of simplicity underlying the third-factor hypothesis, which

Despite the above-mentioned overlap with the contextualist theses propounded by de Regt, Ludlow's argument for this 'ease of use' thesis is unconvincing, we find. This is in large part because it rests on a false dichotomy: namely, that simplicity must be conceived of either as an objective, "absolute" and universal property of theories (possibly complemented by a realist metaphysical justification about the simplicity of reality); or as an always subjective, relative, strictly pragmatic connotation of those theories that allow us to "accomplish our goals with the minimal amount of cognitive labor" (2011, p. 152).

In a sense, we might charitably say that Ludlow's account stops short at de Regt's intelligibility condition; indeed, on the few occasions in which Ludlow mentions understanding (e.g.: "the clearest sense we can make of [simplicity] is [...] in terms of 'simple to use and understand'" (2011, p. 152)) it is reasonably clear that he has in mind what de Regt terms 'understanding a theory.' The merit of the latter's account is that it explores the connection between such pragmatic considerations and the wider explanatory aims and achievements of science. By contrast, as noted above Barrios does acknowledge such connections, both with respect to linguistic inquiry and to science at large. For instance, Barrios offers a reconstruction of generative history which—not unlike the reconstruction presented in our Sect. 2—emphasises the parallelism between the changing role of simplicity on the one hand, and the goals of linguistic inquiry (observational adequacy, descriptive adequacy, explanatory adequacy, explanatory depth) on the other; he also offers an orthogonal analysis that identifies some of the traditional interpretations of simplicity (unification, parsimony) as underlying specific stages of linguistic theory.

Without entering into a detailed discussion of Barrios's rich analysis of simplicity throughout generative history—much of which we agree with—here we merely comment on the main difference between that proposal and the present one. In a nutshell, the divergence stems from our respective conceptions of the aims of scientific (and linguistic) inquiry, as well as of the methods deployed to achieve such aims. As to the former, Barrios seems on the whole to side with a more orthodox conception according to which science (and therefore linguistics) aims at the truth, or some reasonably close proxy. Similarly, Barrios entertains a more or less traditional conception of the vehicles of scientific inquiry, that construes the latter class as exhausted by theories in the standard sense. In contrast, the proposals we are aligning ourselves with support a conception of scientific vehicle that is both more flexible—the relevance of which will become clearer in Sect. 6—and (therefore) more faithful to actual scientific practice. In sum, in both these respects we part ways with Barrios over much the same concerns that separate current accounts of scientific understanding from the more traditional analyses of this notion.

We submit that the perspectives on theory-simplicity presented in this section have potentially significant repercussions for linguistic inquiry. In the next section we finally

Footnote 38 continued

we discussed in Sect. 4—albeit not in terms of reduction. In large part, this is because the term 'reduction' is very rarely employed by generativists, who have indeed occasionally explicitly rejected this interpretation of their practice. But we do *de facto* discuss reduction (albeit horizontal—to domain-general cognitive principles—rather than vertical—to low-level processes) in the context of discussing the third-factor hypothesis (that is, where simplicity becomes an explanatory goal).

put the pieces together, and sketch what we see as a promising research agenda for generative linguistics, philosophy and cognitive science.

6 Everybody wins: talking points for future dialogue

Up until now, we have discussed language acquisition and evolution as largely separate problems. But the two share an important connection, insofar as their respective generative solutions pull in opposite directions: acquisition requires rich, innate linguistic content, and evolution requires a thin, deflated UG. This tension is defused, however, in light of the proposal sketched in Sect. 4: that is, if we set aside the idea that ‘*something* must be special to language’, and countenance the hypothesis that language acquisition could be explained in terms of second- and third-factor content alone. Indeed, we maintain this would qualify as an appealing approach by *minimalist standards*, for several reasons: (1) current empirical research suggests that any ‘solution’ to Plato’s Problem would feature simplicity (as a general cognitive principle) among its main explanatory factors; (2) the hypothesis of a cognitive simplicity principle seems to breathe new life into the early generative insight (Chomsky 1965) that some sort of internal simplicity criterion participates in language acquisition;³⁹ (3) by subsuming language acquisition under a broader cognitive account, (a) the resulting explanation would meet several theoretical desiderata such as coherence, unification and, of course, simplicity; (b) the account would also meet both kinds of naturalist standards—ours, and the minimalist’s (cf. Sect. 3). These reasons are further compounded by a fourth: namely, that the integration of minimalist inquiry into cognitive science would allow for a *unified* treatment of both Plato’s and Darwin’s Problems.

To reiterate, we think that while the foregoing does require a perspective shift on the minimalist’s part, it can still be reconciled with the spirit of (at least some) minimalist tenets. At the beginning of Sect. 3, we remarked on the fluctuations in the interpretation of (both object- and theory-) simplicity between and even within competing frameworks. In fact, diachronic analyses such as ours reveal a subtler trend than this, especially where object-simplicity is concerned. That is, over and above any and all local variations, what remains fixed is the idea that *object-simplicity is language-specific*. Our proposal would require this idea to be revisited rather than abandoned: specifically, to shift from thinking of FL as intrinsically simple (perhaps as a corollary of a sweeping generalisation about the simplicity of nature), to thinking that FL *inherits its simplicity from domain-general features of our cognitive system*.

Indeed, we’re making a broadly parallel point about theory-simplicity. What transpired from Sects. 2–3 is that as a result of their commitment to a hard-nosed realism combined with the Galilean style, minimalists have come to hold an unnecessarily narrow perspective on the available ‘meta’-explanatory options. Among other things, this means that truth (or approximate truth, representational accuracy, etc.) stands unchallenged as the do or die of any one account, at the expense of other epistemic benefits. Here, too, our proposal is of a hermeneutic rather than revolutionary stripe. We’re not suggesting that minimalists toss out any (much less all) of the theoretical

³⁹ See also Yang (2017).

achievements accrued so far. In fact, we're urging that minimalists themselves avoid doing so: rather than holding theoretical products to a single uncompromising standard of truth, other explanatory and epistemic benefits, sanctioned by successful sciences, should be considered.

In addition, it seems to us that the foregoing dovetails very nicely with the philosophical analyses of the role of aesthetic values described in Sect. 5.2. On the one hand de Regt's contextualist account offers an illuminating interpretative key on the fluctuating conceptualisation of simplicity in the course of generative history. Furthermore, both de Regt and yet more explicitly Breitenbach ascribe a more prominent role to theoretical values—including simplicity—in scientific practice, as a result of carving out the relation between scientist-theory-world in a novel way. A third point of contact is seen most clearly by noting a salient difference between the two accounts: while de Regt's main concern is to elucidate the ways in which theoretical values contribute to scientific understanding, Breitenbach is more interested in where these values 'come from'. And, once her proposal is stripped of its Kantian overtones, what remains is a cognitive hypothesis: namely, that aesthetic judgments are the result of the subject's cognitive makeup, and of the interaction between the latter and the world, via theory.

In light of these observations, a few interesting projects suggest themselves. First, we think it would be a fruitful minimalist exercise to examine past and current linguistic practice by the lights of the above philosophical accounts. There are many ways one could implement this somewhat vague suggestion. In what follows we sketch just one of these.

In Sect. 5, we made a point of emphasizing the pluralist orientation of the debate on understanding; this is witnessed, for instance, by the gradual broadening of accepted construals of the notion of explanation, to encompass even mutually incompatible conceptions. Of particular interest is the manifestation of such pluralist tendencies with respect to the *vehicles* of scientific understanding. We've seen this to be a varied class (Sect. 5.1); even more so when we take into account the heterogeneity of its proper subclasses. Indeed the single most diverse of these subclasses is also the most resourced by working scientists: namely the class of scientific models, minimally construed as (more or less idealized) representations of a target phenomenon. That models come in many shapes and forms is well known; for instance, two models about a same target phenomenon P may differ in terms of the degree of abstraction incorporated in their respective representations of P . Models can be highly realistic and concrete (e.g. scale models) or highly idealized and abstract (e.g. toy models). Most interestingly for our purposes, even models that sit at the latter end of the spectrum—that is, even models that are highly simple, idealized and literally false of their target, known as toy models—are widely recognized to be vehicles of scientific understanding.⁴⁰

⁴⁰ A well known example is the Schelling model of racial segregation (Weisberg 2013). The model's target is the phenomenon of segregation in urban areas; its representation of this phenomenon is highly simplified (in that it makes only very few assumptions about the target) and idealized (in that its main assumptions contain deliberate distortions, such as the absence of difference-making socioeconomic factors). Since its inception, Schelling's model (also known as the checkerboard model) has been widely used by social scientists as well as coopted by philosophers to study and illuminate previously undetected features of segregation (and segregation-like) phenomena.

In what way do models so far removed from reality produce, or advance, understanding of their target phenomena? The widely accepted answer is that they do so *precisely as a result of* their deliberate suspension and/or distortion of explanatory factors. More generally, it is (also) in virtue of their extreme simplicity that toy models throw light on phenomena that are either too complex to study directly, or where it is still unclear which factors are genuinely explanatory, and so on. Thus, even toy models are qualified to deliver understanding: specifically, as argued for instance by Reutlinger et al. (2017), they (can) provide a *potential explanation* of their target phenomenon, as a result of which they (can) produce or enhance *how-possibly* understanding of the phenomenon in question.

We think that the foregoing—and more generally, the broader debate on ways in which different epistemic vehicles can function as gateways to scientific understanding—could lead to powerful new insights within generative practice; conversely, we think that generative linguistics should be included in the philosophical conversation on the aims and methods of science. In order to implement this idea, a first and prerequisite step must be for the generative community to liberalize their extant conception of epistemic vehicle, in particular to encompass those which do not satisfy a strict factivity clause (e.g. idealized models). A subsequent key step would then be to reinterpret specific generative theories and hypotheses—attributing ever-increasing simplicity to FL—as candidate vehicles of one or more kinds of understanding.

As a prime illustration, consider P&P. As we saw in Sect. 2, P&P retained a lasting influence (up to and including the early years of MP) insofar as it offered a simple and attractive answer to Plato's Problem, in terms of a relatively small number of abstract, universal, innate principles together with parameters that are switched on or off in response to environmental linguistic stimuli. What went wrong? The standard answer is that P&P is incompatible with evolutionary theory. But another way of seeing things is that P&P was judged (and therefore eventually discarded) *qua* purportedly veridical theory. However, once we liberalize the working conception of epistemic vehicle, new options open up. In particular, it becomes very natural to reinterpret P&P as a highly idealized model of FL: one that suspends at least one explanatory factor (the acquisition process, which is relegated to an infallible on/off switch) and distorts others (the bulk of the explanatory burden is borne by innate, domain-specific content). Once these substantial idealizations are acknowledged, it becomes quite clear that P&P, while implausible and indeed unviable as a veridical theory, can however yield understanding in the form of a potential explanation of its target phenomenon. Thus, P&P helps shed light on questions such as: How much of the explanatory burden of language acquisition can be pushed onto innate, language-specific content? And: Which among the acknowledged explanatory factors (innate linguistic content, acquisition process, primary linguistic data) are genuine difference-makers? And so on. An immediate upshot is then that P&P needn't be discarded *just because* it is false of the actual world. It should rather be judged on its merits *as a vehicle of understanding* of language acquisition.⁴¹

⁴¹ Indeed, we think a case can be made to the effect that P&P yields understanding not just of a modal variety—as suggested here—but also of both heuristic and pedagogical ones (cf. Reutlinger et al. 2017). We are developing both ideas in preparation for a separate article.

In closing, we mention just two more promising angles of future inquiry. First, we think that generative debates hold deep philosophical interest, whereas they have been largely ignored by mainstream philosophy. In particular, we hope to have shown that generative linguistics makes for an intriguing case study on the relation between criteria of scientific understanding, explanatory adequacy, and different interpretations of simplicity.⁴²

Finally, it would be an interesting project to examine Breitenbach's hypothesis itself from an empirical perspective, and more specifically to investigate (i) the cognitive underpinnings of understanding, and (ii) the connection between the latter and the cognitive simplicity principle.

7 Conclusion

This paper started with the observation that, given the centrality of simplicity in their most recent research program, minimalists ought to address the issues of justification and convergence as a matter of urgency. We then outlined and defended a naturalistic approach to both questions; crucially, the proposals outlined in Sects. 4 and 5–6 are accompanied by robust *justifications* of, respectively, the hypothesis that simplicity is a property of FL (insofar as it is a general cognitive principle that interacts with FL to produce domain-specific effects) and the adoption of simplicity as a theoretical value (insofar as simplicity, along with other aesthetic values, is conducive to understanding).

Just as importantly, the proposed account offers a sharper and more nuanced characterisation of both object- and theory-simplicity that rules out the possibility of further conflation of these notions. Conversely, with these sharpened notions in hand it becomes possible to rigorously assess the minimalist expectation that the two should converge.

Finally, we hope to have shown that embarking on a genuinely collaborative path promises to be a fruitful endeavour for minimalists, philosophers and cognitive scientists alike.

Acknowledgements We are indebted to the members of the 'Epistemology of the LHC' research group for organising the *Simplicities & Complexities* conference and sponsoring the associated essay competition that inspired us to write this paper; to the University of Bonn for hosting the event; and to the audience for their warm welcome and stimulating discussions. This was without a doubt one of the best conferences either of us has ever attended. We also thank the audiences of the V Conference of the Brazilian Society for Analytic Philosophy and of the 11th Principia International Symposium. Two reviewers made a number of helpful suggestions to improve the manuscript, for which we are grateful. Finally, we owe special thanks to Christopher Clarke and Anthony Everett for their intellectual and moral support; and above all to Milena Ivanova for (unwittingly) confirming that we were on to something after all.

Funding This work was partially supported by the Brazilian National Postdoctoral Program (PNPD) of the Comissão de Aperfeiçoamento de Pessoal do Nível Superior/CAPES (grant recipient: M. I. Corbalán).

⁴² E.g. ontological versus syntactic (Baker 2016); anti-quantitative vs. anti-superfluity (Barnes 2000); agnostic vs. atheistic Ockham's razor (Sober 2015); Ockham's razor vs. Ockham's laser (Baron and Tallant 2018); quantitative vs. qualitative parsimony in science and philosophy (Lewis 1973; Jansson and Tallant 2016).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Al-Mutairi, F. R. (2014). *The minimalist program: The nature and plausibility of Chomsky's biolinguistics* (Vol. 143). CUP.
- Ankeny, R., Chang, H., Boumans, M., & Boon, M. (2011). Introduction: Philosophy of science in practice. *European Journal for Philosophy of Science*, 1(3), 303.
- Bailer-Jones, D. (1997). *Scientific models: A cognitive approach with an application in astrophysics*. PhD thesis, University of Cambridge.
- Baker, A. (2003). Quantitative parsimony and explanatory power. *The British Journal for the Philosophy of Science*, 54(2), 245–259.
- Baker, A. (2016). Simplicity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (winter ed.). Stanford: Metaphysics Research Lab, Stanford University.
- Barnes, E. C. (2000). Ockham's razor and the anti-superfluity principle. *Erkenntnis*, 53(3), 353–374.
- Baron, S., & Tallant, J. (2018). Do not revise Ockham's razor without necessity. *Philosophy and Phenomenological Research*, 96(3), 596–619.
- Barrios, E. (2016). Simple is not easy. *Synthese*, 193(7), 2261–2305.
- Boeckx, C. (2006). *Linguistic minimalism: Origins, concepts, methods, and aims*. Oxford: OUP.
- Boeckx, C. (2009). The nature of merge: Consequences for language, mind, and biology. In M. Piattelli-Palmarini, J. Uriagereka, & P. Salaburu (Eds.), *Of minds and language: A dialogue with Noam Chomsky in the Basque Country* (pp. 44–57). Oxford: OUP.
- Boeckx, C. (2010). Linguistic minimalism. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis* (pp. 485–505). Oxford: OUP.
- Boeckx, C. (2014a). *Elementary Syntactic Structures: Prospects of a Feature-Free Syntax*. Cambridge Studies in Linguistics. Cambridge University Press.
- Boeckx, C. (2014b). What principles and parameters got wrong. In M. C. Picallo (Ed.), *Linguistic Variation in the Minimalist Framework*. Oxford: OUP.
- Boeckx, C. (2016). Considerations pertaining to the nature of logodiversity. In *Rethinking parameters* (pp. 64–104). Oxford University Press.
- Bonawitz, E. B., & Lombrozo, T. (2012). Occam's rattle: Children's use of simplicity and probability to constrain inference. *Developmental Psychology*, 48(4), 1156.
- Breitenbach, A. (2013). Aesthetics in science: A Kantian proposal. In *Proceedings of the Aristotelian Society* (Vol. 113, pp. 83–100). Wiley Online Library.
- Chater, N., & Christiansen, M. H. (2010). Language evolution as cultural evolution: How language is shaped by the brain. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(5), 623–628.
- Chater, N., Clark, A., Goldsmith, J. A., & Perfors, A. (2015). *Empiricism and Language Learnability*. Oxford: OUP.
- Chomsky, N. (1951). Morphophonemics of Modern Hebrew. MA thesis, University of Pennsylvania, New York.
- Chomsky, N. (1957). *Syntactic Structure*. Mouton.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1975/1955). *The logical structure of linguistic theory*. University of Chicago Press.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Kluwer.
- Chomsky, N. (1993). *Lectures on government and binding: The Pisa lectures*. Number 9. Walter de Gruyter.
- Chomsky, N. (1995). *The Minimalist Program*. Current Studies in Linguistics 28.
- Chomsky, N. (1996) *Powers and Prospects: Reflections on Human Nature and the Social Order*. Pluto Press.

- Chomsky, N. (2004). *The generative enterprise revisited: Discussions with Riny Huybregts, Henk van Riemsdijk, Naoki Fukui and Mihoko Zushi*. Berlin: Mouton de Gruyter.
- Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36(1), 1–22.
- Chomsky, N. (2007). Approaching UG from below. In U. Sauerland & H.-M. Gärtner (Eds.), *Interfaces + recursion = language? Chomsky's Minimalism and the View from Syntax-Semantics* (pp. 1–29). Berlin: Mouton de Gruyter.
- Chomsky, N. (2009). *Cartesian linguistics: A chapter in the history of rationalist thought*. Cambridge: CUP.
- Christiansen, M. H., Reali, F., & Chater, N. (2006). The Baldwin effect works for functional, but not arbitrary, features of language. In *The Evolution of Language* (pp. 27–34). World Scientific.
- Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology*, 6, 1964.
- de Regt, H. (2009). Intelligibility and scientific understanding. In H. de Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific Understanding: Philosophical Perspectives*. Pittsburgh: University of Pittsburgh Press.
- de Regt, H. W. (2017). *Understanding scientific understanding*. Oxford: OUP.
- Dellsén, F. (2017). Understanding without justification or belief. *Ratio*, 30(3), 239–254.
- Di Sciullo, A. M., & Boeckx, C. (2011). *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty* (Vol. 1). Oxford: OUP.
- Douglas, H. (2013). The value of cognitive values. *Philosophy of Science*, 80(5), 796–806.
- Elgin, C. (2007). Understanding and the facts. *Philosophical Studies*, 132, 33–42.
- Elgin, C. Z. (2017). Exemplification in understanding. In Grimm, S., Baumberger, C., Ammon, S. (Eds.) *Explaining understanding: New perspectives from epistemology and philosophy of science*. New York: Routledge.
- Feldman, J. (2003). The simplicity principle in human concept learning. *Current Directions in Psychological Science*, 12(6), 227–232.
- Feldman, J. (2016). The simplicity principle in perception and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 7(5), 330–340.
- Forster, M., & Sober, E. (1994). How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *The British Journal for the Philosophy of Science*, 45(1), 1–35.
- Freidin, R., & Lasnik, H. (2011). Some roots of minimalism in generative grammar. In Boeckx, C. (Ed.) *The Oxford handbook of linguistic minimalism* (pp. 1–26).
- Genewein, T., & Braun, D. A. (2014). Occam's razor in sensorimotor learning. *Proceedings of the Royal Society B: Biological Sciences*, 281(1783), 2013–2952.
- Gershman, S., & Niv, Y. (2013). Perceptual estimation obeys Occam's razor. *Frontiers in Psychology*, 4, 623.
- Gijssbers, V. (2013). Understanding, explanation, and unification. *Studies in History and Philosophy of Science*, 44, 516–522.
- Grimm, S. R. (2006). Is understanding a species of knowledge? *The British Journal for the Philosophy of Science*, 57(3), 515–535.
- Grimm, S. R., Baumberger, C., & Ammon, S. (2017). *Explaining understanding: New perspectives from epistemology and philosophy of science*. Abingdon: Routledge.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298(5598), 1569–1579.
- Hills, A. (2016). Understanding why. *Noûs*, 49(2), 661–688.
- Hornstein, N., Nunes, J., & Grohmann, K. K. (2005). *Understanding minimalism*. Cambridge: CUP.
- Hsu, A. S., Chater, N., & Vitányi, P. (2013). Language learning from positive evidence, reconsidered: A simplicity-based approach. *Topics in Cognitive Science*, 5(1), 35–55.
- Ivanova, M. (2017). Aesthetic values in science. *Philosophy Compass*, 12(10), 1.
- Jansson, L., & Tallant, J. (2016). Quantitative parsimony: Probably for the better. *The British Journal for the Philosophy of Science*, 68(3), 781–803.
- Janvid, M. (2018). Getting a grasp of the grasping involved in understanding. *Acta Analytica*, 33(3), 371–383.
- Kertész, A. (2010). From 'scientific revolution' to 'unscientific revolution': an analysis of approaches to the history of generative linguistics. *Language Sciences*, 32(5), 507–527.
- Khalifa, K. (2013). Is understanding explanatory or objectual? *Synthese*, 190(6), 1153–1171.
- Khalifa, K. (2017). *Understanding, explanation, and scientific knowledge*. Cambridge: CUP.
- Kinsella, A. R. (2009). *Language evolution and syntactic theory* (Vol. 1). Cambridge: Cambridge University Press.

- Kitcher, P. (2019). So ?who is your audience? *European Journal for Philosophy of Science*, 9(1), 1–15.
- Kosso, P. (2002). The omniscienter: Beauty and scientific understanding. *International Studies in the Philosophy of Science*, 16(1), 39–48.
- Kvanvig, J. (2003). *The value of knowledge and the pursuit of understanding*. Cambridge: CUP.
- Kvanvig, J. (2009). The value of understanding. In A. Haddock, A. Millar, & D. Pritchard (Eds.), *Epistemic value* (pp. 95–112). Oxford: Oxford University Press.
- Lappin, S., Levine, R. D., & Johnson, D. E. (2000). The structure of unscientific revolutions. *Natural Language & Linguistic Theory* (pp. 665–671).
- Lawler, I. (2019). Scientific understanding and felicitous legitimate falsehoods. *Synthese* (pp. 1–29).
- Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
- Li, M., & Vitányi, P. (1997). *An introduction to Kolmogorov complexity and its applications*. Berlin: Springer.
- Lipton, P. (2009). *Understanding without explanation*. In *Scientific Understanding: Philosophical Perspectives*. University of Pittsburgh Press.
- Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences*, 20(10), 748–759.
- Ludlow, P. (2011). *The Philosophy of Generative Linguistics*. Oxford: OUP.
- Narain, D., Smeets, J. B. J., Mamassian, P., Brenner, E., & van Beers, R. J. (2014). Structure learning and the Occam's razor principle: a new view of human function acquisition. *Frontiers in Computational Neuroscience*, 8, 121.
- Nersessian, N. J., (ed.) (1987). *The Process of Science: Contemporary Philosophical Approaches to Understanding Scientific Practice*. Number 3 in Science and Philosophy. Kluwer Academic Publishers.
- Onnis, L., Roberts, M., & Chater, N. (2002). Simplicity: A cure for overgeneralizations in language acquisition? In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 24).
- Pothos, E. M., & Chater, N. (2005). Unsupervised categorization and category learning. *The Quarterly Journal of Experimental Psychology*, 58(4), 733–752.
- Reutlinger, A., Hangleiter, D., & Hartmann, S. (2017). Understanding (with) toy models. *The British Journal for the Philosophy of Science*, 69(4), 1069–1099.
- Schindler, S. (2018). *Theoretical Virtues in Science: Uncovering Reality Through Theory*. Cambridge: CUP.
- Smith, N., & Allott, N. (2016). *Chomsky: Ideas and ideals*. Cambridge: CUP.
- Sober, E. (1975). *Simplicity*. Oxford: Clarendon Press.
- Sober, E. (1978). Computability and cognition. *Synthese*, 39(3), 383–399.
- Sober, E. (2002). What is the problem of simplicity? In H. A. Keuzenkamp, M. McAlleer, & A. Zellner (Eds.), *Simplicity, inference and modelling*. Cambridge: CUP.
- Sober, E. (2015). *Ockham's razors: a user's manual*. Cambridge: CUP.
- Soler, L., Zwart, S., Lynch, M., & Israel-Jost, V., (eds.) (2014). *Science after the Practice Turn in the Philosophy, History, and Social Studies of Science*. Routledge Studies in the Philosophy of Science 14. Routledge.
- Strevens, M. (2016). How idealizations provide understanding. In *Explaining understanding: New perspectives from epistemology and philosophy of science*. Routledge.
- Tallerman, M., & Gibson, K. R., (eds.) (2012). *The Oxford handbook of language evolution*. Oxford: OUP.
- Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences of the United States of America*, 113(16), 4530–4535.
- Van Fraassen, B. C. (1980). *The Scientific Image*. Oxford: OUP.
- Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford: OUP.
- Yang, C. (2017). Rage against the machine: Evaluation metrics in the 21st century. *Language Acquisition*, 24(2), 100–125.
- Zagzebski, L. (2001). Recovering understanding. In M. Steup (Ed.), *Knowledge, truth, and duty: Essays on epistemic justification, responsibility, and virtue*. Oxford: Oxford University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.