



Empiricism, syntax, and ontogeny

Gabe Dupre

To cite this article: Gabe Dupre (2021): Empiricism, syntax, and ontogeny, *Philosophical Psychology*, DOI: [10.1080/09515089.2021.1937591](https://doi.org/10.1080/09515089.2021.1937591)

To link to this article: <https://doi.org/10.1080/09515089.2021.1937591>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 14 Jun 2021.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Empiricism, syntax, and ontogeny

Gabe Dupre 

School of Social, Political, and Global Studies, Keele University

ABSTRACT

Generative grammarians typically advocate for a *rationalist* understanding of language acquisition, according to which the structure of a developed language faculty reflects innate guidance rather than environmental influence. This proposal is developed in developmental linguistics by *triggering* models of language acquisition. Opposing this tradition, various theorists have advocated for *empiricist* views of language acquisition, according to which the structure of a developed linguistic competence reflects the linguistic environment in which this competence developed. On this picture, linguistic development is accounted for by general statistical learning mechanisms. In this article I shall precisify the debate, provide a clearer picture of what is at stake, and show why an intermediate picture is needed.

ARTICLE HISTORY

Received 31 May 2019
Accepted 27 May 2021

KEYWORDS

Philosophy of psychology;
philosophy of linguistics;
generative grammar;
developmental linguistics;
rationalism/nativism;
empiricism; bayesianism

1. Rationalism and empiricism

The terms ‘rationalism’ and ‘empiricism’ have been applied to many distinct positions. Traditional rationalists¹ and empiricists² adopted clusters of views relating epistemology, metaphysics, and psychology, and followers of these programs extended them to other areas.³ I will be concerned with how this debate plays out in psychology, wherein modern rationalists claim that psychological states and capacities are innately determined, whereas empiricists claim development is a process of reflecting the environment.

This raises the question: what does it mean for psychological states to be innately determined? Like ‘rationalism’ and ‘empiricism’, ‘innate’ has been used to draw many non-equivalent distinctions. For this reason, Griffiths (2002) argues that we should eschew this concept entirely in favor of more specific ones. This strikes me as an over-reaction. As long as we are careful to specify exactly how we are using the term, and to interpret other’s claims in line with their stipulated usage, the harms of equivocation can be avoided. In a series of papers Mameli and Bateson (2011, 2006); Bateson and Mameli (2007), have argued similarly that the folk notion of innateness is used to pick out a wide range of different properties, many of which

CONTACT Gabe Dupre  G.g.dupre@keele.ac.uk

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

are scientifically uninteresting, and have thus urged that we may be better off eschewing this notion entirely, in favor of scientifically better-defined notions such as canalization or adaptation. Again, I believe that, as long as one is explicit about how the term is being used, these problems can be avoided. I am not here attempting to provide an analysis of the term 'innate'. I am interested instead using the term to draw a distinction between certain classes of models of psychological development, focusing on language acquisition.

As I shall use the term, a trait is innate to the extent that an explanation of its structure does not require reference to the extraction of information from the developmental environment. Crucially, this is not merely a *causal* claim. All developed states are causally dependent on both environmental and internal/biological influences, and so drawing this distinction along causal lines is a nonstarter.⁴ Likewise, the claim is not that innate traits don't carry information about the environment. In the mathematical sense of 'information', any reliable causal dependency will result in the carrying of information.⁵ The motivation for this view is that some traits seem to be molded by the environment, reflecting the structure of environmental stimuli, whereas others, even though causally dependent on the environment, develop structures that do not reflect the properties of the environment. In such cases, the structure of the trait must be internally given. The rationalist claims that aspects of the mind are structured by internal forces and thus do not, or at least need not, reflect the properties of the environment. The empiricist, on the other hand, claims that developed psychologies are reflective of patterns in the environment to which organisms are sensitive.

This proposal is thus intended to be consistent with approaches to development which claim to reject a strict distinction between biology and environment, such as Developmental Systems Theory (as developed in Oyama et al. (2003)). These approaches stress the interdependence between organism and environment, undermining the factorization of the explanation of developed traits into those that are "in the genes" and those that are acquired. They also stress the ways in which whatever role genetics does play is only made possible by substantial scaffolding from non-genetic (both organismic and environmental) structures. As I read it, the impetus behind such proposals is that the causal relations between organism and environment are complex and dialectical. I do not deny this, and I am not assuming that the traits I describe as 'innate' need be specified or determined by the organism's genome.⁶ What I am stressing is that, even given this causal complexity, we are able to distinguish between two kinds of development: one which relies on the environment purely causally, and another which responds to patterns in this environment by functioning to reproduce these patterns within the organism.

This account also allows for the appropriate, if etymologically puzzling, claim that innate traits need not be present at birth. An innate trait can develop at any time, as long as the structure of the end product is not reflective of any external stimulus.⁷

It also has the nice result that it makes sense to talk of a trait as being more or less innate. A trait is entirely innate if *all* of its structure is a result of internal forces, even if it depends causally on external stimuli. A trait can also be entirely non-innate if its structure exactly matches that of an environmental pattern to which the organism is sensitive. Traits which partially reflect the environment, but deviate from the environmental patterns in internally driven ways will count as partially innate. This account is thus not subject to the worry that the debate is empty because all traits are causally dependent on internal and external factors. Rationalists must accept that the environment plays a role, and empiricists must accept that minds respond to the environment in specific and contingent ways. However, the rationalist claims that the role of the environment is limited to influencing which of a limited set of options the system will be internally driven to develop, whereas the empiricist views the mind as a system for attuning an organism to its environment, with development involving coming to more accurately reflect this environment. The crucial distinction, as will become clear when we look at triggering models of language acquisition, is between, on the one hand, the environment *selecting* or *enabling* the development of a trait, and on the other, the environment providing the details for how the trait is supposed to be structured.

One final feature of my proposal is that we may wish to adopt a *relativized* or domain-specific notion of innateness, according to which features of some model of acquisition may be innate in an unrestricted sense, i.e., influence the development of a trait in ways other than (rationally) reflecting the environment, but will not be counted as innate from the perspective of the developmental task in question. I have in mind here features of development which deviate from pure reflection of the environment, but in ways which are common throughout developmental systems. For example, it may be a general feature of learning that learning mechanisms instantiate certain biases, e.g., preferences for relatively regular rules. This would show that learning is not purely empiricist, but would not show that there are any innate features of traits which are specific to the trait in question. If language acquisition depended only on these very general biases, there would be a clear sense in which language was not innate. So we can distinguish between purely empiricist systems in the widest sense, which incorporate no mechanism for deviating from the observations, and purely empiricist systems in the narrow sense which incorporate no mechanism for such deviations specific to the areas in which they apply. We will look in section 4

at a particular acquisition model which should hopefully make these distinctions clear.

Here is a simple example exemplifying the contrasts between rationalist and empiricist positions. Imagine two systems responding to strings of letters. One system, the empiricist system E, uses statistical mechanisms to abstract patterns from the input string. The other, the rationalist system R, instead produces a repeating string of the first letter of the input. Compare E and R's responses to the following input strings:

- (1) aaaaaa . . .
- (2) ababab . . .
- (3) all the king's horses

For string 1, E and R would produce the same output. R, however, would produce exactly the same response to all three input strings, whereas E would differentiate string 2 from string 1, accurately reproducing both of them. For 3, however, E would either produce something nonsensical or nothing at all.⁸ There are a few things to notice here. Firstly, E is much less restricted in its output states. E has as many possible states as there are patterns detectable by its algorithm. R however is highly restricted. It can produce only 26 representations. Relatedly, while both R and E's final states are caused by their environments, only E's are reflective of the environment. From E's final state, we can predict E's environment with a decent degree of accuracy. However, for R this is not the case. An output of 'aaaa' is consistent between any number of inputs. This is the sense in which non-innate traits carry more information about the environment. Likewise, if R does resemble the environment, this is, from a developmental perspective, an *accident*: it just so happens that of the many possible environmental patterns R could have encountered, it found one of the few it can resemble. This is not the case for E. Hopefully the picture of innateness just described can be made a little clearer through comparison to some of the other standard accounts in the literature, which I shall turn to in the next section.

2. Developing the account

An account close to my own in the literature is the view that innateness should be viewed as canalization, as defended by Ariew (1996, 1999) and Collins (2005). A trait is canalized to the extent that its properties are independent of the variation in the environments in which the trait develops.⁹ Roughly, we can think of canalized traits as those for which there is a many-to-one mapping of developmental environments onto traits and their properties. For example, the human skeleton, if it is able to develop at all, acquires roughly the same structure in whichever environment it

arises, and so is highly canalized. On the other hand, what sort of music one likes seems to depend greatly on the fine-grained properties of one's environment, and so is not canalized. This theory then says that the degree to which a trait is innate is the degree to which it is canalized: innate traits develop broadly independently of their environments, while non-innate traits are highly sensitive to the properties of the environment.

My own account overlaps significantly with the canalization approach. Given that a canalized trait, by definition, develops in the same way in many different contexts, explanations of the developed structure of such traits are unlikely to require much appeal to the environment. If, in a wide range of environments, the same trait develops, it is highly likely that the developed structure of this trait is explained with reference to internal features of the organism. Thus, canalized traits are likely to be counted as innate on my theory as well. However, the converse does not hold. Traits which develop only in a specific set of environments *may* nonetheless incorporate internally driven structure, and so will, to that extent, count as innate on my theory. This reflects the fact that certain developmental pathways may require very specific environmental stimuli to develop, but these stimuli are nonetheless mere causes of such development and do not structure the developed trait.

As with the canalization account of innateness, I believe that my proposal can incorporate the insights behind *primitivist* accounts of innateness (see e.g., Cowie (1999) and Samuels (2002, 2004) without also adopting their pitfalls. Primitivist accounts view a trait as innate if and only if its development is not a result of *psychological* processes. This account nicely captures the sense in which Fodor (1975, 1981, 1998) views concepts as innate: i.e., they are not acquired as a result of *learning* or some other rational process. However, as Collins (2005) points out, this account fails to meet an apparent desideratum for a theory of innateness in that it applies only to psychological traits. Obviously, we cannot apply this theory to traits more generally, as it would lead to non-psychological processes of organismic change, such as getting a tattoo, counting as innate. Of course, it could be that there are simply distinct notions of innateness in the cognitive and biological sciences, but all things being equal an account of innateness which applied to both biological and psychological development would be preferable.

I believe Collins' objection points to a broader issue: there is no reason, in principle, to assume that the distinction between non-psychological and psychological processes tracks the distinction between processes of reflecting or being structured by the environment and processes which are merely caused by the environment. If I am right, the notion of innateness, as exemplified in the linguistic disputes I will be discussing, aims to capture the latter, rather than the former. Indeed, there is reason to think these

distinctions may cross-cut one another. Triggering models of language acquisition, to be discussed later, are fully fledged psychological processes: they posit the development of a particular system of linguistic rules on the basis of representation of linguistic input. But these are nativist theories in that the structure of the developed system is guided by, and thus explained with reference to, internal properties of the human language faculty, not statistical features of the linguistic environment. In the other direction, we may view certain developmental processes as functioning to replicate features of the environment by way of non-psychological processes. Burge (2010) (Chapter 9) presents an extended argument that certain sensory systems, especially gustation and olfaction, should not be viewed as genuinely perceptual systems, on the grounds that such capacities can be explained without the positing of genuine psychological *representations*. If we take representation to be a genuine *mark of the mental*, then we could read this work as denying that such capacities are genuinely psychological. But we would not want to infer that all influence of such systems, e.g., on the behavior of non-perceptual organisms such as amoebae or worms, or in humans prior to their integration with genuinely representational processes such as generalization and categorization, is innate. This suggests the possibility of non-psychological processes of extraction of information from the environment, which would count as non-innate on my view, but not that of the primitivist. This seems like the right result. Of course, the philosophical and empirical premises here are controversial, but it seems that a theory of innateness should leave open the possibility of such environment driven, but non-psychological, development.

This set of considerations also tells against the proposal in Fodor (2008) to replace the innate/acquired distinction with the distinction between states acquired brute-causally and those acquired rationally, while again holding onto the insight behind this distinction. Rationality is conceptually, I take it, a constraint on representational processes. But as we saw above, there seems to be no reason in principle to think that the *only* way for an organism to develop so as to reproduce the properties of the environment is by representing the environment. While rational processes will, I suppose, be processes of this sort, “brute causal” processes need not fail to reproduce the properties of their causes. But when such brute-causal processes do lead to the development of traits which reflect the environment, intuitively these traits will not be plausibly viewed as innate.

Margolis and Laurence (2013) argue that the nativism/empiricism debate should be conceived of as a debate concerning the *number* of psychological mechanisms: empiricists view all developed (psychological) traits as stemming from the operation of a small number of generally applicable mechanisms, while the nativist views these traits as a product of many specialized systems. I believe that *extensionally* my view and that of Margolis and

Laurence are fairly similar. But I believe my account is better as a description of what is actually at stake. The question is: *why* should we care so much about how many mechanisms must be appealed to in explaining development? My account provides an answer: if the development of a trait depends on a specialized system, the character of that system is likely to play an essential role in the explanation of the structure of the developed trait. On the empiricist view, different domains of knowledge (language, mathematics, social reasoning, etc.) are differentiated only by the information from which they have been generated: the mechanism for their acquisition is the same, general-purpose, learning system. So, in explaining why these domains of knowledge are as they are, we will appeal only to these differences in the input, i.e., the environment. For the nativist/rationalist on the other hand, if each of these domains is a product of distinct acquisition mechanisms, we will likely have to appeal to these distinct mechanisms in accounting for their developed properties. This means that we can replace Margolis and Laurence's initially puzzling claim that the debate is *characterized* by the issue of how many mechanisms there are, to the claim that the number of mechanisms is a characteristic point of disagreement between nativists and rationalists, and this disagreement stems from the explanatory strategies of these respective approaches.

O'Neill (2015) argues that our notion of innateness should be relativized. In particular, she claims that a trait is innate *relative to some particular* environmental attribute when its development is insensitive to variation in that attribute. So for example, having a working visual system is innate relative to the language spoken in one's environment, but not relative to one's eyes being sewn shut for the first few months of life (Wiesel and Hubel (1963)). While I agree that the innateness of a trait should be relativized, I disagree with O'Neill's approach on the grounds that it retains the problems of any account of innateness based on the causal role of the environment. As stated earlier, the problem with this is that all traits are causally dependent on the environment in some ways. This is less of a problem for O'Neill than for some other views, as it means that every trait will only be non-innate *relative to some environmental factor*, rather than categorically non-innate. But this still seems to blur the key distinction between different types of developmental process as detailed in this paper. For example, all traits will presumably, on O'Neill's account, be non-innate relative to nutrition. If an organism cannot acquire enough energy, it simply won't develop at all. Likewise, all parties to the debate about language acquisition agree that some linguistic stimulus is required in order for language to develop.¹⁰ This dependence on the environment seems crucially different from the dependence on the environment appealed to in empiricist theories.

As my account appeals to the explanation of developed traits, it is worth touching on some important features of explanation in general which will be

important to my account. Firstly, explanations typically discriminate between causes. While many things are required to cause some event, only some subset of these will be relevant to explaining it. While the spark and the oxygen are needed to *cause* the fire, only the spark (usually) explains it. Secondly, explanations are contrastive: one and the same event can receive different explanations when contrasted to different sets of alternatives. While consumption of alcohol can explain why one got a tattoo, it can't explain why someone got a tattoo *of the phrase 'Mexican Death Star'*. For that, we need more fine-grained information about the circumstances (see Lipton (1990) for classical discussion of both of these points). Both of these points will then apply to the classification of a trait as innate. While encountering language in the environment may be necessary for the development of language, nativists claim that this experience does not (typically) play a central role in explaining language acquisition. Just as the presence of environmental oxygen is backgrounded in explanations of why the house burned down, experience of some language can, according to these theorists, be largely backgrounded in explanations of language acquisition. Likewise, how we characterize the trait at issue will influence what the appropriate explanation will be. That I speak *English*, rather than Korean, will be explained with reference to my being raised in an English-speaking environment (although whether the *structure* of English is explained in this way is precisely the issue to be discussed later). But that I speak *a language with a hierarchical syntax* will not.

On the face of it, my notion of innateness is susceptible to some of the stock counter-examples in the literature. For example, what Damasio (1994) calls 'acquired sociopathy' (p. 178), anti-social behavioral disorders caused by neurological damage, will count as innate, as the resultant states and behavior are not reflective of their environmental cause. Clearly, the environmental cause, e.g., a blow to the head, does not here provide the structural details of the acquired psychopathology. So, contrary to our intuitions, acquired psychopathy will thus be counted as innate on my approach.

In such cases, the explanation for why certain kinds of trauma lead to these anti-social states will involve primarily an account of the nature of the internal processes which lead, in response to this trauma, to the resultant state. The mind is, in this respect, unlike a car's bumper. A physical shock to the bumper may create a dent with roughly similar size and shape to the object responsible. But, the dynamics of the brain mean that gross physical damage leads to significant endogenous restructuring. This restructuring can thus have surprising psychological and behavioral effects which do not, in any way, reflect properties of the cause of the damage. As these effects are quite systematic between agents (James and Blair (2002)), it seems that an explanation of them must appeal to predictable internal features of

organization and plasticity. To this extent, it seems to not do too much damage to our notion of innateness to view this traits as innate.

Note that the counter-intuitiveness of this result can be lessened even more by the contrastive nature of explanation, and thus of innateness. If we want to explain why someone has acquired sociopathy vs. not acquiring sociopathy, that they have received a blow to the head will be of central explanatory interest. Thus, *that* they are sociopathic is not innate: explanations of this trait must appeal to the environment. However, that sociopathy has the characteristic behavioral and psychological properties that it does will instead contrast this outcome to other possible responses to trauma. If we ask why this person became impulsive, aggressive, insensitive to the legitimate moral claims of others, etc., rather than, say, becoming placid and generous, we can hold the environmental stimulus fixed and explain this with reference to internal features of the victim's brain. To this extent, then, (this aspect of) the trait will be innate. The trait will thus be innate in the people in which it occurs, as its structure is explained without reference to the environment, but the occurrence or not of the trait is not innate. The environmental stimulus, the cause of trauma, plays a purely causal role in explaining the acquired psychological state.

Likewise, Fodor (1975, p. 32) introduced the thought experiment of a "Latin pill" which, when ingested, would grant one the ability to speak Latin. As is usual with such sci-fi examples, what they show depends on details which are underspecified in their description. Perhaps surprisingly, the "mechanism" of such a pill will determine whether the acquisition of Latin in this way will count as innate or not. One possibility would be that the pill contains, in some microscopic format, all the structure of Latin, and somehow this structure is extracted by the brain of the person ingesting the pill. This would then be a case, as Fodor intended, of a trait which is neither innate nor learned. But it poses no problem for my view, as intuitively in such a case we are extracting information from the environment, albeit in a surprisingly compressed form, and so would be non-innate as predicted. Alternatively, the pill could serve to directly modify features of the mind/brain which are modified in the normal course of acquisition by environmental linguistic data. One possibility here would be that the pill serves as a sort of simultaneous trigger to the various parametric linguistic options made available to the human mind. In such a case, it seems intuitively reasonable to call the acquisition of Latin 'innate', as the structure of the developed language is again explained with reference to the internal states of the mind. In a sense, the subject knew Latin all along, and the pill just served to make this knowledge available. The case which is needed, to provide a problem for this view, is one in which there is no suitable structure present in either the pill or the pre-ingestion mind, but where the pill causes the mind to acquire Latin nonetheless. However, such a case seems like either

inexplicable random chance or magic. The structure of Latin, with all its declensions and conjugations, must come from somewhere! If not, it is far from clear that there could ever be a systematic and law-like explanation of such a causal link. As what is being proposed is a theory of innateness suitable for empirical science, it seems that such examples can be safely ignored.

Before moving on to the empirical questions that such an account makes possible, it is worth explaining why I do not take it to be essential that innate traits are adaptations, as does Lorenz (1966). On the face of it, such a condition can help resolve certain puzzles about innateness, on my view, including the apparently counter-intuitive results just discussed. As humans have (presumably?) not been selected for their ability to become sociopathic after taking a blow to the head, this trait would be ruled out by this condition. Thus, something like Fodor's three-way distinction between innate, learned, and neither innate nor learned could be recapitulated, with the final category (applying to *inter alia* acquired sociopathy) containing states which are explained without reference to extraction of information about environmental structure but which are not adaptations.¹¹ Despite this benefit, I do not want it to be a requirement that innate traits be adaptations. For one reason, as Barrett (2014) discusses at length, natural selection plays an essential role in explaining *both* innate and non-innate traits. Evolutionary forces have led to organisms with traits with a wide variety of norms of reactions, roughly functions from environment and genotype to phenotype. Even learned traits are made possible by selection for the capacity to learn in these ways. So, being a product of adaptation will not distinguish innate from non-innate traits. In the other direction, and perhaps more significantly for my purposes, whether our linguistic capacities are adaptations is itself highly controversial. Chomsky has consistently argued over the years that we ought be skeptical of adaptationist arguments in this domain (dating back at least to Chomsky (1972)), and recent work in the Minimalist Program has argued that it may be plausible that much of the structure of the language faculty stems not from contingent adaptations but instead from so-called 'third factors', general constraints on the development of complex systems (Chomsky (2005)). This would thus be an innate trait which is not adaptive. While this idea is highly controversial in the area of language, since the work of Gould and Lewontin (1979) it is uncontroversial that some traits will be explained in this manner. So it should not be a constraint on innateness that it be adaptive.

I hope to have shown that my account differs, in favorable ways, from the variety of options currently available in the literature. My discussion of these views has, of course, been quite quick, and has not captured all the subtle differences between versions of these various views. But I hope it has at least served to further elucidate my account of innateness as absence

of extraction of environmental information, and shown what makes it plausible.

3. The structure of the arguments

Turning now to the linguistic case, a developmental theory of language aims to identify the function from available linguistic data to acquired language faculty. We can assume as a starting point that we have access to the input/output patterns: what state the developed linguistic competence¹² adopts in response to what primary linguistic data. The goal then is to work backwards from this mapping to discover the processes at work. Of course, the temporal order of linguistic theorizing is not this simple. Linguists do not need to wait until the final, correct grammatical theory has been proposed before developing theories of acquisition, and often which grammatical theory is deemed best will depend on how competing theories interact with developmental theories. We can, however, adopt the idealization that developmental theories are proposed to explain the accepted input-output patterns for the purposes of this paper.

As in the case of our machines, there are characteristic patterns of input/output behavior which provide evidence for the competing positions. Patterns supporting rationalism will show:

1 a) Information in output (e.g., developed linguistic competence) not found in input (e.g., primary linguistic data).

b) Few possible output states.

c) Low resemblance between input and output.

Empiricist-favoring patterns however will show:

2 a) All information in output found in input.

b) Many possible output states.

c) High resemblance between input and output.

Pattern 1a is the clearest case for rationalism. If there is information in the developed system not found in the stimulus, this must have been contributed by the learner.

Patterns 2b and 2c constitute (less decisive) evidence for empiricism. If we observe a high degree of resemblance between the environment and the developed states of the learner, this *could* be a result of a highly structured initial state which happens to correlate with the environment in which it developed. However, the closer the correlation is, the less parsimonious such a proposal will be. This proposal would rely on the claim that significant correlation between the organism and the environment is, from an ontogenetic perspective, *accidental*. An empiricist proposal would not need to posit some extra mechanism to explain (away) any organism/environment correlation. This is just what such systems are designed to do, and so empiricist accounts are neater when such correlations are observed.

Note that there are plausible rationalist accounts of why, despite the correlation between internal state and external environment, the state should be viewed as innate. The most general such account would be an evolutionary one. If environmental patterns are highly stable, it is conceivable that over evolutionary time the reproduction of these patterns would be offloaded from ontogeny to phylogeny.¹³ That is, if it is adaptive to know about certain features of one's environment, and these features remain the same over evolutionary time-scales, organisms with this knowledge innately may acquire an adaptive benefit, in that they need not spend resources extracting this knowledge from the environment, and so minds containing such knowledge may simply develop, rather than needing to learn it each generation. In such a case, I would view the reflection of the environment as innate.¹⁴ To distinguish such a case of innate structure from empiricist, acquired structure, we would need to, perhaps counterfactually, see what happens when the environment ceases to be reliable. The nativist system would continue to develop as it normally does (assuming that the environment is not so catastrophically different that development is impossible), and thus cease to reflect the environment, whereas the empiricist system would develop differently, reflecting the novel properties of the new environment. Much work in Evolutionary Psychology (e.g., Barkow et al. (1995)) assumes this sort of picture, including the claim that organisms, especially humans, may be in trouble because our innately developing minds are structured by an environment that is no longer present. Of course, this is not to defend this research, and there are many criticisms of both this evolutionary proposal and the developmental processes it presupposes,¹⁵ but it is at least a possibility. This further indicates the complexity of inferring from these observations which *prima facie* suggest empiricist, or rationalist, models.

Likewise, pattern 2b is *consistent* with a rationalist proposal. The initial state could be highly complex, allowing for external stimuli to select one of a large number of possible resultant states. However, if the structure of the system's final states is fully provided innately, then as the number of possible final states increases, so does the burden placed on innate structure. If we observe significant variation in the final state, this variation is likely to be more economically accounted for by empiricist mechanisms which extract this information from the environment.

We can now identify the various positions in debates about linguistic innateness. A pure rationalist claims that all the structure of the adult competence is provided by internal constraints. This view allows for language variation, but only in the sense that different environmental stimuli can cause the faculty to develop along different lines. But the stimuli are *merely* causal. The developed faculty need not resemble or reflect the properties of this cause in any substantial way. At the other extreme,

a pure empiricism will claim that *all* of human linguistic competence reflects, at some level, properties of its linguistic environment. Chater et al. (2015) propose, in this vein, that language acquisition can be viewed as a process of *lossless compression*: the aim of the learner is to find the maximally compact way of describing the most likely source of the linguistic data they encounter.¹⁶ For example, if a language contains constraints on wh-movement, then such constraints must be evidenced in the input.¹⁷

Intermediate positions involve the claims that *some* language learning is rationalist or empiricist. The weak rationalist claims that at least some of the structure of the developed language faculty is not provided by the environment. The weak empiricist claims that at least some of this structure is. Of course, unlike the pure positions, these positions are compatible. Indeed, if ‘some’ is read to mean *some but not all*, these positions will be mutually entailing. If the languages one learns partially reflect innate structures and partially reflect one’s linguistic environment, weak versions of both positions will be vindicated.

As mentioned in section 1, this account allows for a *graded* notion of innateness. This is, I think, a virtue, as many, if not most, traits seem partially reflective of the environment, and partially structured by internal factors. I shall argue that linguistic competence should be viewed in exactly this way, as a mixture of both internal and external factors. In general, this notion of innateness should enable us to tease out exactly which aspects of a particular learning model are innate and which are not.

It will, I am sure, be clear that many idealizations are being made in this description of the goals of developmental linguistics. Most centrally, the idealized picture I have presented takes for granted that the inputs to the process of language acquisition, the primary linguistic data, and the outputs, the developed competences, are known. Of course, much of the debate turns precisely on these issues. As a matter of sociological fact, views on the input, output, and process cluster together. Rationalists tend to take the input to be sparse, the output to be complex, abstract, and categorical, and the process to be correspondingly in need of substantive innate guidance. Empiricists dispute the poverty of the input (see e.g., Reali and Christiansen (2005)), have less abstract understandings of the acquired competence, and thus feel less need to posit substantive innate constraints.

For example, one’s view on what shape developed linguistic competence takes will motivate different sorts of learning system. Mainstream generativist theories, such as Minimalism (Chomsky (1995)), have mostly viewed linguistic rules as *categorical*: a syntactic construction is either generable by the system or it is not.¹⁸ This doesn’t rule out a probabilistic acquisition process, and indeed both early (Chomsky (1965)) and recent (Yang (2002, 2016)) work in generative linguistics suggests such a model, but it does open the door for a categorical model of learning, as we shall see in discussion of

triggering models. On the other hand, if one views our linguistic competence as itself consisting in probabilistic information (see e.g., Manning (2003) and Chater et al. (2015)), specifying not merely which expressions can be combined, but how *likely* such combinations are, then this seems to require a model of acquisition which keeps track of statistical patterns in encountered language. Relatedly, different views of what is involved in knowing a language, i.e., whether it is knowing a grammar as in mainstream generative theories, or developing a collection of autonomous linguistic modules (syntactic, semantic, phonological) as in Jackendoff (2002) and Culicover & Jackendoff (Culicover & Jackendoff, 2005), or maintaining a store of constructions as Tomasello (2003) and Goldberg (2006) urge, or even simply in connections between expressions (McClelland & Patterson (2002)) will likewise fundamentally influence what sort of learning model will seem appropriate.¹⁹

Additionally, the categories of rationalist and empiricist models, or aspects of models, are themselves very coarse grained, blurring over important distinctions between kinds of model. Of course, the maximally simple imitation represented by my toy empiricist model above is woefully less powerful than any actual empiricist proposal in the literature. And such proposals will vary widely in the assumptions they make and the ways these assumptions guide their extrapolation from data. Any model based on imitation will trivially not be able to differentiate between linguistic structures which are licensed but not attested in the data from those which are not attested because they are illegitimate. Much of the work in empiricist theorizing aims precisely to account for this, and thus overcome the Poverty of Stimulus argument (more on this later). For example, Bayesian approaches (e.g., Perfors et al. (2011) and Chater et al. (2015)) are capable, in some circumstances, of utilizing the likelihood that a given expression-type *would be* absent from the data if it were licit to determine whether it is likely absent from the data as a result of a prohibition or chance (But see fn. 27 for worries with this approach). The debate between the high-level categories of rationalism and empiricism thus must involve much more fine-grained comparison of specific rationalist and empiricist models alongside detailed understanding of the data available to the learner and precise theories of the acquired competence.

Further, even if there were agreement on the *actual* primary linguistic data and the linguistic competences actually acquired on this basis, this does not suffice to determine the function used to map the former onto the latter. Empiricist and rationalist models of acquisition may agree on the observed cases but diverge in unobserved situations. This comes across in disputes about the richness of the data. When rationalists claim that some linguistic rule/constraint cannot be learned from the primary linguistic data, empiricists often respond by showing that there is some evidence for this rule in

linguistic corpora.²⁰ The issue then becomes whether such evidence is sufficiently frequent not only for a child exposed to these data to acquire the rule in question without substantive linguistic biases, but also whether it is frequent enough to ensure that all other speakers of the language were themselves exposed to analogous data, and if not whether they indeed acquire the rule/constraint in question. Given these uncertainties surrounding the input-output structure of the learning problem, disputes between theorists cannot take place strictly at the computational level, but will instead be sensitive to questions at the lower levels of algorithm and implementation (Marr (1982)). That is, our assumptions about what the inputs and outputs of the learning process *would be* in unobserved cases can, at least in principle, be guided by the differential plausibility of the mechanisms (algorithmic and neurobiological) needed to implement distinct computational level theories which may equally provide solutions to the mapping problem consistent with our observations of actual language learners and their environment.²¹

As always, theory choice here must be holistic. We cannot really start with a model of the available data, or of the acquired competence, as a *fixed point*, and then mold our developmental theories accordingly. What is assessed is thus a complex of theories/models of all three domains (and more, including neuroscience, other branches of cognitive science, evolutionary theory, etc.). Having noted this, I hope it will not distort the discussion too much to continue to assume a relatively fixed picture of the primary linguistic data and of the developed language faculty. The aim here of course is not to establish a particular developmental theory, but instead to display how such debates function, to show that there are no *a priori* answers to be had, and tentatively suggest that an intermediate approach, weak rationalism and weak empiricism, is most plausible given the current state of play.

4. An example: Bayesian grammar acquisition

Culbertson and Smolensky (2012) present a Bayesian model of acquisition of word order rules for a small fraction of English, consisting of just three basic expression types: N(ouns), Adj(ectives), and Num(eral determiners). The aim of the model is to determine, on the basis of some observed expressions, whether ‘modifiers’ (Adj and Num) are grammatical before or after the nouns they combine with. The grammars in question are probabilistic, rather than deterministic, so technically all four options (Adj-N, N-Adj, Num-N, N-Num) are grammatical, and what needs to be determined are the probabilities of each. That is, some grammars make, for example, Adj-N more likely than N-Adj, and the model aims to determine which such grammar is most likely responsible for the data it encounters.

A grammar can, in this model, be defined as a pair of conditional probabilities: p_{Adj} : (Adj – N|AdjP), and p_{Num} : (Num – N|NumP): these express the probability of the modifier preceding the noun, given that a modifier of this sort is present. “(Adj – N|AdjP) = 1” means that adjectives always precede nouns, for example. The model estimates, on the basis of encountering some training data (a collection of modifier/noun pairs with an equal number of pairs containing Adj and containing Num, but with different proportions of modifiers before or after the noun), which pair of such conditional probabilities is most likely to have produced these data. This is the grammar that is then inferred to.

Along with the training data, the priors of the system determine which grammar is inferred to (in line with Bayesian updating). There are two factors of the priors in these models:

A: A *regularization bias*.

B: A *substantive bias*.

The regularization bias is the degree to which extreme values (close to 1 or 0) of p_{Adj} and p_{Num} are preferred. The regularization bias is encoded in the model as the shape of the prior probability distribution over grammars. This is given as the combination of four components. Each component corresponds to a preference for an equivalence class of grammars, namely those that favor a particular pair of word orderings. For example, one component will favor a grammar which views Adj-N and Num-N orders as most likely (although members of this class will assign different probabilities to these orderings, they will all be greater than .5). The other components likewise favor (Adj-N, N-Num), (N-Adj, N-Num), and (N-Adj, Num-N) grammars. An important restriction in this model is that the *degree* to which these extreme values are favored is the same in all four components. This degree is given as the shape of the probability distribution each component assigns to its favored grammar, given as a product of two beta distributions: $(\alpha_{Adj}, \beta_{Adj})$ representing the shape of the prior probability distribution for hypotheses concerning the frequency of Adj-N vs. N-Adj, $(\alpha_{Num}, \beta_{Num})$ doing likewise for Num-N vs. N-Num. The ratio of $\alpha:\beta$ is determined by the model.

The substantive bias involves assigning weights to each component, increasing or decreasing the regularization in the direction of the component’s favored grammar. This is stated as a set of four values, one for each equivalence class of grammars $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$. These values always sum to 1, but the particular values are determined by the model.

The priors of the model are thus given by the four beta distributions, each imposing the same degree of regularization toward the four possible classes of grammars, and four component strengths, which each serve to increase or reduce the influence of one of these four components. Differences in the values of these priors influence the behavior of these systems in different ways:

- (1) Models with flat beta distributions (i.e., with $\alpha=\beta=1$) do not regularize, and so will adopt the grammar that precisely tracks the distribution of word orders in the training data.
- (2) The more uneven the distribution (i.e., the further the ratio of $\alpha:\beta$ deviates from 1), the less variability in the data will be reflected by the selected grammar. Distributions with high α values will regularize toward grammars with pre-nominal modifiers, while distributions with high β values regularize toward grammars with post-nominal modifiers.
- (3) The values of γ influence which equivalence class of grammar is favored. For example, a high γ_3 increases the regularization toward (Num-N, N-Adj) grammars.

Model training involves running Bayesian inferences on the training data, for a collection of models with different priors (i.e., different regularization and substantive biases). Each model will select the grammar which would make the training data most likely. Model testing then involves introducing new data, and seeing which of the grammars arrived at by each model, with their distinct biases/priors, makes these testing data most likely. Those priors which resulted in the model best able to predict the training data are then viewed as correct. In this case, a strongly regularizing prior, biased toward (Adj-N, Num-N) and (N-Adj, N-Num) grammars, with low weight for (N-Adj, Num-N), and zero weight for (Adj-N, N-Num) maximized the likelihood of the training data. This distribution is in line with Greenberg's Universal 18, which says that Adj-N languages are usually Num-N languages, but not conversely.

This model demonstrates the various contributions of the environment and the learning system I wish to classify. In the first place, this is a weakly empiricist model. The Bayesian response to the environmental data aims to extract enough information from these data to determine what was the most likely source of these patterns. Clearly, different environmental data lead to different acquired grammars, and the selected grammar reflects patterns in the data. However, the priors, both the substantive and regularization bias, contribute essentially to the selection of a grammar. The regularization bias means that the acquired grammar *does not* merely reproduce the statistical patterns in the environment, but instead assumes that such a pattern was noisily generated, and so projects a simpler rule. The substantive bias means that some of these simpler (regular) rules are favored over others, with the extreme case being (Adj-N, N-Num) toward which the model never regularizes. The extent to which the acquired grammar deviates from the environmental pattern, due to these priors, is thus the extent to which language is innate according to this model.

It is worth noting also that the two biases, which together determine the shape of the priors, seem to differ with respect to how plausibly general they are. It is likely that some sort of noise-reduction mechanism is widespread in psychological learning, and so the regularization bias may plausibly be viewed as innate in the wider sense, but not innate with respect to language learning. If this were the only bias in a language acquisition model, it would be fair to describe language as non-innate. However, the substantive bias is language specific. It is stated as a preference for some word orders over others. It is hard to see how this could reflect a general property of inductive inference. For this reason, according to this model, language acquisition is, at least to some extent, innate in the domain-specific sense. This seems to be the intuitive description of this model: it is explicitly designed to capture the ways in which human speakers are disposed to favor certain specifically linguistic hypotheses over others.

While this is an impure empiricist model, it indicates what a pure model would look like. As I said, removing the substantive bias plausibly would produce a model which is, from the perspective of language, purely empiricist: the developed state can be explained without any reference to language-specific innate influence. While we cannot remove the priors entirely, we could remove the bias inherent in them. This could be done by positing a flat distribution, removing the preference for regular grammars. Such a system would simply reproduce the statistical patterns in the environment: if 80% of the observed sentences containing adjectives were Adj-N, the system would infer that 80% of all adjectives preceded their nouns.²²

The explanatory power of this model essentially depends on the priors it assigns to the learner. We can explain Greenberg's universal 18 with reference to the substantive bias. This is in contrast with one of the often stated virtues of Bayesian models: that the priors *don't* matter. The more explanatory work is done by the priors, the less is done by the Bayesian confirmation itself, and so it is often argued that the explanatory power of certain Bayesian systems does not depend at all on the priors. This is because it is provable that under a variety of conditions, given sufficient evidence, the priors *wash out*. That is, as the evidence increases, posteriors will become arbitrarily close to one another, no matter what priors they started with. This points to another distinction helpfully captured by my notion of nativism: while all Bayesian models must have priors, they only sometimes play an explanatory role in accounting for development. When the priors wash out, we do not need to make reference to them in explaining the developed trait, and thus can view such a trait as non-innate. However, when, as in the case above, the shape of the priors plays an essential explanatory role, the developed trait is, to that extent, innate.

As well as general worries with Bayesian models of language acquisition (See e.g., Yang (2017)), there are worries specific to this model. In particular,

it is highly restricted in scope, focusing on the acquisition only of the word ordering of three kinds of phrases. Of course, starting with the simpler cases is often a good methodology, but it is not always possible to scale up. In this case, one difficulty is that the tested hypotheses (the conditional probabilities of modifiers preceding/following nouns) cover the entire possible hypothesis space: all orders, with all probabilities, are tested. One of the central insights of the generativist tradition is that this will not be the case in general for language acquisition: some linguistic hypotheses are never even entertained, such as hypotheses positing structure-independent rules. Relatedly, it is unlikely that hypotheses as specific as those in this model are tested. Even if there are word-order rules involved in acquiring a language, they are likely more general than those in the Bayesian model discussed. Determining which hypotheses are included and excluded will thus in general be a further area relevant for claims about linguistic nativism. I discuss this model, despite these difficulties, just to provide a worked-out case in which we can clearly apply the various distinctions about innate and non-innate aspects of learning models. In particular, I hope that this model has shown that the question of whether a system is purely empiricist, purely nativist, or a mixed system is not *a priori*. Given a suitable definition of innateness, all of these are empirical possibilities. I turn now to the argument that language acquisition must in fact be mixed.

5. Against pure empiricism

The central argument for rationalism, the Poverty of Stimulus argument (hereafter PoS), is precisely an argument for the claim that linguistic competence is innate in the sense earlier defined. The logic of a PoS involves determining that some aspect of a developed language faculty (e.g., constraints on movement) is present even when there is nothing in the environment (the stimulus) for it to reflect. Under these conditions, this extra structure must be provided by the organism itself, and is thus innate.

The logic of PoS is as follows²³:

P1: In acquiring their native language, learners adopt hypothesis H rather than distinct hypothesis H'.

P2: The evidence available does not discriminate between H and H'.

C: Therefore, a preference for H is innate.

It is usually granted that such an argument is (barring magic) valid. If the evidence available to the child does not discriminate between the two hypotheses, there must be some innate fact about the child which does. The case for innateness then depends on finding linguistic hypotheses that are acquired, but for which favorable evidence is not plausibly found in the environment.²⁵

One paradigmatic example is yes-no question formation. Consider the relationship between the indicative (1) and the corresponding interrogative (2):

- (1) Xian has gone to Panama.
- (2) Has Xian gone to Panama?

English-speaking children by the age of three are uniformly able to form and understand questions in this way. The problem is that such examples of question-formation, which are plentiful in a child's linguistic environment, do not settle which rule is used in question formation. For example, the rule "front the first auxiliary" would correctly predict this pattern.²⁶ However, consider the following, more complex, sentences:

- (3) Xian, who has been fired, has gone to Panama.
- (4) Has Xian, who has been fired, gone to Panama?
- (5) *Has Xian, who been fired, has gone to Panama?

The rule just proposed would predict that (5) would be the interrogative form of (3). But (5) is nonsense, as every English-learning child knows. The rule in question must rather be "front the matrix auxiliary", which correctly predicts sentence (4). Premise 1 of the argument is thus established.

The argument is completed by showing that the evidence requisite for discriminating these two rules (e.g., sentence (4)) is not available to the language-learning child. Legate and Yang (2002) calculate that of the 20,651 questions in the CHILDES corpus, only 14 (0.068%) are relevant to selecting the matrix auxiliary rule over the first auxiliary rule.²⁷

As all children learn the correct rule, and never make mistakes like sentence (5), in order to deny premise 2 of the argument the evidence available must be robust enough that we can assume every child will encounter it in sufficient quantity. Such low frequency suggests that this is not the case for data like sentence (4).²⁸ So, the argument concludes, because the child is able to select a linguistic hypothesis even if she does not have the requisite evidence favoring this hypothesis over others, the preference for hypotheses of this kind must be innate.

PoS, however, establishes only weak rationalism. Such arguments show that *some* structure for the adult language system is provided by innate constraints rather than abstracted from the environment. While this argument is not universally accepted, its logic is good and the kinds of cases discussed above, *inter alia*, have yet to be sufficiently addressed by the strong empiricist. I shall therefore assume that weak rationalism at least is correct. The next stage of the debate involves working out whether more can be

claimed for the rationalist picture. Is there reason to say that *all* linguistic competence is provided in this way?

6. The possibility of pure rationalism

Pure rationalism about language is the view that the human ability to acquire a language is very similar to that of R's ability to represent strings.²⁹ A small set of possible final states of the language faculty are circumscribed by its initial state, which also specifies which of these final states will result when confronted with what primary linguistic data.

The central challenge for such an approach is accounting for predictable linguistic variation, the apparent exemplification of patterns 2b and 2c above. While PoS arguments establish an important role for innate/internal forces in structuring the adult competence, *prima facie* the learner's linguistic environment plays a similar role. This accounts for the basic fact that *ceteris paribus* speakers raised in England learn to speak English, and speakers raised in Japan learn to speak Japanese. That is, just as with E above, properties of the environment are predictable on the basis of properties of the developed competence. As mentioned, this is not a conclusive argument for empiricism, but is strong *prima facie* evidence for it.

A defense of pure rationalism must thus account for both linguistic diversity (2b) and the apparent reflection of the environment by developed language faculties (2c). The Principles and Parameters approach attempts to do just this. According to this approach, we can divide the constraints on the human language faculty into two types. Principles are absolute constraints on what languages humans can learn. During acquisition, languages violating these principles are not even considered. As well as principles there are parameters, which provide a *range* of possible options. Natural languages may differ in which of these options they select. The traditional picture views parameters as principle-schemas: providing the form of a rule that a language must follow, but including an open slot which must be filled to determine the exact content of the rule. For example, we can think of the Head Directionality Parameter as stating: "Heads must be -- their complements", where ' -- ' is filled in by the child, in response to linguistic experience, with either 'before' or 'after'. By reducing language variation to the setting of these parameters, it is claimed that much of the apparent diversity is a surface-level phenomenon underlain by deep similarities.³⁰

The pure rationalist attempt to account for the apparent reflection of the environment is provided by triggering models of language acquisition.³¹ According to these accounts, language acquisition (especially the setting of parameters), involves deterministically adopting one aspect of a grammar on the basis of a relatively small exposure to linguistic data. For example, the Null-subject parameter (determining whether a sentential subject must be

pronounced or not) may be triggered by exposure to just one or a handful of sentences without pronounced subjects. The environment, according to such a model, is causally significant in determining the adopted grammar, but need not be reflected by the grammar. That is, there need not be any “rational” relation between the trigger and the result triggered. R, above, is an example of a triggering system.³²

The general picture is captured nicely in Chomsky (2000, p. 8): “*We can think of the initial state of the faculty of language as a fixed network connected to a switch box; the network is constituted of the principles of language, while the switches are the options to be determined by experience. When the switches are set one way, we have Swahili; when they are set another way, we have Japanese. Each possible human language is identified as a particular setting of the switches – a setting of parameters, in technical terminology. If the research program succeeds, we should be able literally to deduce Swahili from one choice of settings, Japanese from another, and so on through the languages that humans can acquire.*”

On such a picture, human languages are more similar than they initially appear. They all exist within the narrow possibility space provided by principles, and surface variation is a result of different parameter settings. In particular, a small(-ish) collection of parameters may, given complex cascading interactions between them, result in superficially radically different languages (see e.g., Baker (2002) for a proposal of this sort). As these parameter settings are triggered by, rather than extrapolated from, the environment, all of their structure is given by the innately specified language faculty. This is radically unlike the pure empiricist proposal which takes surface variation at face value and accounts for it by treating language learners as mirroring their divergent environments.

The triggering account of parameter-setting is highly controversial due, in large part, to the problem of ambiguous triggers.³³ For many parameters, there are very few, if any, sentences in the primary linguistic data which unequivocally show how a parameter must be set. For example, differences in surface word-order can result from either differences in base-generation or in movement rules.³⁴ This means that the learning algorithm must use complicated procedures to determine which stimuli are to count as triggers so as to avoid setting a parameter incorrectly.

Much of the work in defending triggering approaches involves providing some story as to how this problem is solved. The Null-subject parameter provides a clear example. English sets this parameter negatively, while Spanish sets it positively, hence the difference between **“am hungry”* and *“tengo hambre”*. This parameter determines whether subjects *can* be dropped, not whether they must: *“Yo tengo hambre”* is perfectly grammatical. English sentences, with pronounced subjects, are therefore consistent with both settings of this parameter. Gibson and Wexler (1994) provide

a solution to this problem, *default parameters*. For various parameters, unambiguous evidence is available only for one parameter setting. By assuming that one parameter is a default, and the other parameter will only be selected given unambiguous evidence for it, the problem of parameter setting becomes tractable. Spanish speakers, confronted with sentences lacking explicit subjects will be forced to accept one setting of the Null-subject parameter, while English speakers will not be given such unambiguous evidence and will thus remain in the default, negative setting.³⁵

The important thing to notice is the asymmetry between arguments for weak and pure rationalism. Whereas in the case for weak rationalism, the argument aims to show that alternative accounts are *impossible*, the argument for strong rationalism involves producing complex models to show that such an account is *possible*. This mirrors the difference between the observation of pattern 1a, which *entails* a contribution from the organism,³⁶ and that of 2b and 2c, which *suggest* an empiricist algorithm, but which are consistent with pure rationalism.

The empiricist picture is thus motivated by arguing that rationalist proposals, like the triggering accounts just discussed, are implausible. Correlation between the environment and the developed linguistic competence is, in light of these proposals, insufficient on its own. What must be shown is that a theory which accounts for this correlation by appealing to antecedently available information internal to the system is less attractive than one which treats this information as originally located in the environment and then reproduced in the psychology by some rational process.³⁷ This can involve claims that positing internal direction leads to bad predictions about development (e.g., the learner is predicted to acquire a rule not reflected in behavior), or that such posits are explanatorily surplus (e.g., the mechanisms needed to extract this information are already required in some other aspect of the psychology).

Which way the debate about parameter settings turns out will have repercussions in the rationalist/empiricist debate. Triggering accounts like Sakas and Fodor (2012) are currently waning in popularity, compared to more statistical proposals like Yang (2002), but this debate is far from settled. However, there are phenomena that tell more decisively against pure rationalism.

7. Against pure rationalism: Semi-productive rules

Some linguistic generalizations are unrestricted in their application. For example, the rule for forming the English progressive, “add -ing”, applies without exception: every progressive form in English is formed in this way. Because these rules are so homogeneously adhered to in the environment,

such cases provide little discriminating evidence between the empiricist and the rationalist. English speakers' developed states show no variation with respect to this state, and so the rationalist does not need to posit much internal structure to account for this phenomenon. They can also explain why such a phenomenon is so wide spread: it is an aspect of one of the small number of stable states achievable by the language faculty. The empiricists, on the other hand, can similarly explain this with reference to the speaker's reflection of highly stable environmental patterns.

However, some rules are more restricted in their application. Consider dative alternation, wherein the order of the direct and indirect objects of transitive verbs is switched, and the direct object is marked with a preposition.

- (6) I gave Maria the ball.
- (7) I gave the ball to Maria.

Such pairs are semantically equivalent. However, this pattern is not universally applicable. There are cases where the switched (8–9) or unswitched (10–11) order is not available:

- (8) The water gave Maria typhoid.
- (9) *The water gave typhoid to Maria.
- (10) *I sacrificed God a lamb.
- (11) I sacrificed a lamb to God.

While almost every aspect of language acquisition remains mysterious, the acquisition of semi-productive rules is particularly astounding. In the case of the acquisition of the English progressive, a child must recognize that certain utterances describe ongoing events as such. However, once this is done, the child has a clean basis from which to induce a rule. The child can safely assume that verb-stems followed by '-ing' are progressives. Such an extrapolation will not significantly mislead the child in either production or interpretation.³⁸ Compare this to the case of dative alternation. The simplest rule, move the direct object after the indirect object and mark it with 'to', is falsified by (9) and (10). Nevertheless, the child is able to acquire mastery of these rules, and apply them to novel cases in much the way that adults do.³⁹

Yang (2016) discusses a mechanism by which these complex rules are acquired. According to his Tolerance Principle: a rule is acquired if and only if $e \leq \frac{N}{lnN}$, where e is the number of observed exceptions to the rule (i.e., examples to which the rule could apply, but to which it is known not to) and N is the number of examples to which the rule can (correctly or incorrectly) apply. The Tolerance Principle captures a fact about cognitive economy:

extracting generalizations from data is efficient to the extent that the exceptions are few compared to the cases to which the rule correctly applies.

A central feature of Yang's view is the division of data into fine-grained categories for the application of the Tolerance Principle. Based simply on an unclassified natural language corpus, very few rules would meet the threshold required for acquisition. Yang's clearest case of this is German pluralization.⁴⁰ Only about 4% of German nouns found in the corpora pluralize with '-s'. Nonetheless, this rule is productive: novel and loan words typically pluralize with '-s'. How can this be if the exceptions to this rule ($\approx 96\%$ of German nouns) vastly outnumber the good cases? Yang's answer is that the class of constructions to which a rule could apply is not homogenous. Instead, it is divided into multiple sub-classes (e.g., according to gender), each of which are targets of more specific rules. N for each rule is not simply the number of expressions to which the rule *could* apply, but this number minus the number of expressions covered by other rules. In this way, Yang shows how each rule learned ("feminine nouns are pluralized with '-en'" etc.) can meet the threshold required by the Tolerance Constraint. "Add-s" is the most general ("default") rule, in that it doesn't place any special constraints on the nouns to which it applies. The fact that most nouns are already covered by alternative, specific, rules explains why it generalizes: of the few nouns not already covered by other rules, it applies with few exceptions.

Returning to our earlier example, the simplest rule for dative-alternation is: a double-object construction is grammatical if and only if a to-dative is grammatical. However, due to examples like (9) and (10), this won't do. For this reason, Yang argues that speakers taxonomize the class of such constructions so as to find more specific constructions for which the number of exceptions is tolerably low. For example, if we semantically restrict the set of to-datives so as to include only verbs of caused possession (e.g., 'give', 'donate' etc.), the ratio of e to $\frac{N}{\ln N}$ drops sufficiently so as to pass under the bar presented by the Tolerance Principle. Similar subclassification according to semantics, phonology, etc. enables Yang to make accurate predictions about which rules will be adopted, and thus about what kinds of over-generalization will occur.

In this way, when acquiring rules for semi-productive phenomena, the model posits that speakers extrapolate linguistic rules on the basis of how frequently these rules are attested in their environment (relative to observed counter-examples). This is a clear case of an empiricist form of learning, in our sense. The crucial difference between this learning mechanism and triggering proposals is that the latter require that certain stimuli are individually causally sufficient for a given change in the language faculty. Yang's model, however, treats the faculty as responding instead to high-level *patterns*. This empiricist move allows for an explanation of when the faculty

does and does not induce a rule, thus accounting for differential degrees of productivity. For each (possible) class of constructions to which one such rule applies, the pure rationalist would have to posit innate structure. Given how fine-grained these rules are, this would put serious pressure on the innate endowment.

This sensitivity to environmental patterns also explains characteristic properties of linguistic development, such as the “U-shaped curve”: the phenomenon of children’s linguistic competence degrading before they acquire mastery. Very young speakers tend to make fewer grammatical mistakes than those slightly further along in development. Yang explains this with his Tolerance Condition. At a very young age, children are exposed to relatively few verbs, many of which (e.g., ‘be’, ‘have’) are irregular. This prevents extrapolation, and so each inflection is learned separately. As they acquire more verbs, and thus more regular verbs, they begin to project rules, and so we see characteristic over-regularizations (‘I breaked/taked the toy’). As more evidence becomes available, they begin the process of sub-categorization described above, eventuating in mastery of the rules. This model thus makes linguistic competence highly correlated with the specific evidence available, indicative of patterns 2b and 2c above.⁴¹

I hope the preceding has provided a convincing case for an intermediate position between the pure forms of rationalism and empiricism. While PoS shows that *some* information in the adult linguistic competence must be provided by the organism itself, this should not lead one to the stronger claim that natural language acquisition is *purely* internally driven, as generativists have sometimes suggested. The acquisition of semi-regular rules is better explained by a model according to which the learner reflects statistical patterns of usage in the environment in their developed linguistic faculty. This is an important way in which language acquisition is *not* analogous to the development of bodily organs, as Chomsky has often suggested.⁴² While environmental differences may lead to the development of cancer in different organisms’ livers, these effects do not *reflect* their environment. As Yang’s model shows, linguistic competence is, unlike the development of the liver, at least in part a *rational* process, involving the extraction of information in the environment, as the weak empiricist claims.

8. A (Surprising?) convergence

Presumably as a result of the widespread misconception that pure rationalism and pure empiricism are *a priori* false, there is a (perhaps) surprising convergence of views in the literature on the mixed view defended herein. Theorists who are strikingly different in their conceptions of what language is and how it is acquired all propose models in which domain specific innate knowledge provides the space in which empiricist models can apply.

On the one hand, committed empiricist theorists typically model language as hypothesis confirmation, where the hypotheses being tested are themselves provided innately. Following Fodor (1981, 1975), a recent manifesto promoting empiricism in linguistics, Chater et al. (2015, pp. 47–48), accepts that language acquisition must involve searching for the right hypothesis in an antecedently defined search space. As Perfors (2012) says, “a pre-specification of a latent hypothesis space is *necessary* for learning.” (p. 134). But this pre-specified search space is itself precisely an instance of language-specific innate knowledge. Just as we saw in the Bayesian model discussed in section 4, these empiricist models build in language specific information, and then identify which such hypothesis is most plausible in light of the data.

These models thus end up looking a lot like the rationalist models they aim to reject. Early generative work (e.g., Chomsky (1965)) viewed language acquisition as precisely a matter of selecting between alternative possible grammars on the basis of linguistic evidence (as noted by Chater et al. (2015, p. 60). Recent work in this tradition has likewise suggested this model (Yang (2002, 2016)). The major difference, for our purposes, between these views is that the rationalist models in a sense posit *less* innate structure. A robust Universal Grammar consists, in these rationalist models, in the *absence* of certain linguistic hypotheses. Hypotheses describing certain logically possible languages (e.g., languages containing structure-independent rules) are never even entertained by language learners. So the nativist views the language learner as selecting within a constrained search space. But the structure of the problem is the same.

It is worth noting that work in the Minimalist program in contemporary generative theory is likewise moving in this direction. As argued by Boeckx (2010, 2014), building on Chomsky (2002, 2007a, 2007b, 2015), Minimalist motivations make plausible a distinction between the core properties of grammar, the innate and species-universal structure-building operations, and the periphery, consisting of learned and thus variable strategies of *externalization* (i.e., morphology and phonology) and the lexicon. These latter features of language can then be acquired in more-or-less empiricist ways. While this differs from classical generativist proposals in differentiating the learned from the innate as distinct components of acquired linguistic competence (perhaps even reserving the term ‘language’ for the former), we can see a deep similarity in their treatment of language acquisition and variation: innate and universal features of the mind limit the possible developed capacities while systems for extrapolating from linguistic experience account for inter-linguistic variation.

From this perspective, we can see that the major outlier here is the middle-period of the generative tradition, the Principles and Parameters model. Since the downfall of the purely empiricist behaviorism in the early

20th Century, only this view has seriously considered a pure approach to language acquisition, with parameter setting as a deterministic response to particular linguistic stimuli, rather than the kind of inductive approach characteristic of empiricist models. That this model seems to be waning in popularity (see fn. 30) suggests the promise of something of a consensus in the field, although of course this still leaves much room for disagreement.⁴³

9. Conclusion

In this paper, I hope to have shown that pure rationalism and empiricism are untenable. Viable theories of linguistic development will thus need to account both for information provided by the environment and the organism. This position thus differs, on the one hand, from those who claim that pure empiricist (e.g., domain-general Bayesian) or pure rationalist (e.g., strictly triggering-based parametric) theories are sufficient to account for language acquisition, and on the other from those who claim that is is *a priori* that no pure approach will work. By spelling out the logic of the arguments against these strong positions, the strategy of the middle-ground is made clearer, as is its status as a contingent, empirical hypothesis.

Notes

1. E.g. Descartes (1641/2017).
2. E.g., Locke (1836/1996).
3. E.g., the Logical Empiricists' extension from epistemology into scientific methodology.
4. It is sometimes suggested that this fact does not show that innateness is not a causal notion. Instead, it shows that innateness is not a *categorical* notion. That is, that there are degrees of innateness: traits are more or less innate to the extent that they are caused by internal forces. However, as internal and external influences are always necessary but insufficient for developed traits, it is hard to make sense of this graded notion of causality.
5. The carrying of information (i.e., the reducing of possibilities) can however be used as a *heuristic*: innate traits will typically carry less information about the environment than non-innate traits, due precisely to the fact that the latter (function to) reproduce environmental patterns.
6. Although they may be. See Wexler (2003) for an argument, drawing on a very wide range of evidence, that certain features of a child's innate knowledge of language is genetic.
7. Note that strictly, the claim is not that, for a rationalist system, the final state *doesn't* reflect the environment, but that the system does not function to reflect the environment. Criticisms of rationalist proposals often fail due to a misunderstanding of this point. I will often talk simply of traits reflecting or not reflecting the environment, but it should always be kept in mind that what is meant is a functional claim about how this system develops, not merely a feature of the end-product of this development.

8. Obviously, nothing has been said about the details of E, whether it is, say, a Bayesian or a frequentist system, whether it responds to higher-order regularities or not, etc. which would determine its response to such strings. The example is merely illustrative.
9. O'Neill (2015) rightly points out that this is a weaker notion than 'canalization' as introduced into the biological literature by Waddington (1942). Waddington's notion of canalization was a particular *kind* of invariance; namely, invariance in development as a result of developmental feedback mechanisms, usually as a product of selection for such invariance. While Ariew often suggests this more complex notion, most of the literature on innateness (e.g., Samuels (2004) and Collins (2005)) seems instead to view canalization as simple invariance, as stated in the body of the text.
10. So called 'wild children', raised in awful conditions in which they receive minimal linguistic input, never seem to fully acquire a language (see Curtiss et al. (1974)). However, it seems that the linguistic input can apparently be relatively impoverished, as seen in the case of home sign (see Goldin-Meadow (2005)).
11. Note that similar results could be obtained instead by appealing to a "normal environments condition", according to which a trait is not innate if it develops only in abnormal environments, as in Samuels (2002). Such a condition faces all the usual worries with spelling out what it takes for an environment to be normal.
12. One difficulty in describing these debates as neutrally as possible is that empiricists sometimes object to using language like 'competence' or 'faculty' which at least suggest a specialized system. I shall continue using these terms simply to refer to the capacity to use and acquire a language, without presupposing the standard generativist account of in what such capacities consist.
13. A plausible mechanism for this would be the Baldwin effect, wherein genetic dispositions to learn new behaviors are selected for. Glackin (2011) argues for an evolutionary account of language in just these terms.
14. However, Christiansen and Chater (2008) argue in the opposite direction that we may plausibly view the rapidity with which language changes compared to the evolution of the human organism as requiring that it is language which adapts to the requirements of the speakers, and not vice versa. It is unclear what exactly to make of this proposal. Of course, it is agreed by all parties that language must, in order to be passed on from one generation to the next, be *learnable*, and so any "attempt" to modify language which cannot be learned by the next generation will fail. The dispute then is whether the features of the mind which make some languages better adapted are specific to language or not.
15. See e.g., Lewontin et al. (1984), Rose and Rose (2010), Buller (2006), and Lickliter and Honeycutt (2003).
16. Note that pure empiricism is not the claim that no properties of the organism matter for explaining the developed state. This position is indeed *a priori* false. How the organism responds to its environment will of course depend on what the organism is like. A pure empiricist system will have some structural properties which explain why it responds to the environment in a purely empiricist way, namely some developmental system (e.g., a learning algorithm) which functions to precisely reproduce the environmental patterns it encounters.
17. Those objecting to the perceived excesses of generative grammar tend to adopt a position something like this Goldberg (2006), Tomasello (2003), and Onnis et al. (2008)).
18. Hale and Reiss (2008) argue for an analogous view of phonology.

19. I am here assuming that debates about the purview of linguistics (e.g., whether linguistics is, following Chomsky, a science of the mind, or like Katz (1980) a formal theory of Platonic abstracta, or extra-mental concrete reality as in Devitt (2006)) do not arise in developmental linguistics. A theory of language *acquisition* must be a psychological theory, on pain of changing the subject. Of course, however, one's views about linguistic theory may influence one's theory of acquisition.
20. E.g., Sampson (1989) and Pullum and Scholz (2002).
21. Note that I am here assuming that both the rationalist and the empiricist models are themselves computational level models. On the first point, there is wide agreement. However, it is sometimes suggested that many paradigm empiricist models, especially Bayesian approaches, are best understood as algorithmic models. I am here following some of the leading voices in Bayesian cognitive science in insisting that they be viewed, just like their rationalist opposite numbers, as computational (Chater et al. (2011)).
22. But see Van Dongen (2006) for worries that the appeal to flat priors itself hides substantive assumptions which themselves influence the end result in often surprising ways. While van Dongen's point is well taken, his advice on how to avoid bias in Bayesian modeling seems more difficult to adopt. Firstly, he proposes that a certain amount of prior knowledge should be allowed in the selection of priors. This is fine advice in many cases, but doesn't apply to debates concerning nativism, where the shape of the priors may be exactly what is at issue. So, while his point is a sound one, the "uninformative priors approach" may be plausible in this case, even though it is not applicable across the board.
23. See Crain and Pietroski (2001) for an excellent overview.
24. I am assuming here a model according to which language acquisition is profitably described as the acquisition of hypotheses (alternatively, rules or constraints, which I will use interchangeably). This has itself been challenged (see McClelland and Patterson (2002)).
25. As in my above account of reflecting the environment, it is worth distinguishing the functioning of the system from the actual relationship between the system and the environment. In this case, what matters is whether the learner uses or relies on the available evidence to acquire H over H'. If not, H is innate, even if there is (unused) evidence available in the environment. The absence of such evidence from the environment is, of course, the best possible reason for claiming that learners are not using this evidence.
26. I am sidestepping the question of how these abstract classifications (like 'auxiliary' and 'matrix') are themselves acquired by the learner, although this is itself part of a powerful argument for nativism.
27. This is an oversimplification. Crucially, it is highly contentious whether only positive evidence of this sort should be included in the data set from which learners generalize. In particular, Bayesian models, such as Perfors et al. (2010), often stress that the *absence* of certain constructions from the learner's experience can itself function as evidence that such constructions are not possible. There are, however, significant problems with this kind of argument. See, in particular, Marcus (1993) and Yang (2015) for compelling empirical arguments that any system capable of excluding possible expressions on the basis of indirect negative evidence is liable to massively overgenerate and exclude many perfectly acceptable expressions. This problem is exacerbated in cases where the child's language environment is very sparse, as in cases of deaf children with non-signing parents (Goldin-Meadow and Yang (2017)). As mentioned above, the poverty of the linguistic data has itself been challenged by e.g.,

- Realı and Christiansen (2005). Gulordava et al. (2018) develop computational models for the acquisition of hierarchical structure from a naturalistic corpus. See Yang et al. (2017) for critical discussion of these kinds of model.
28. Various features of language acquisition can be used to strengthen this claim. For example, children must not only be *exposed to* these crucial data, they must also *attend to* them. As young children have been shown to be fairly weak at parsing complex sentences, it is not a given that even if they encounter sentences like 4 they will use them as evidence for or against their linguistic hypotheses.
 29. Perhaps even R is slightly impure in that the letter which gets represented as output is always found in the input, as the first letter of the string. R was described this way for clarity, but even this reflection of the environment could be dropped. Say instead that R's behavior was modeled by a set of 26 rules of the form: "if the first character of the input is 'a', output 'bbbb . . .'", "if the first character is 'b', output 'kkkk . . .'", and so on for all 26 classes of possible inputs.
 30. This picture of parameters is somewhat out of date. Contemporary generative theory largely either views parametric variation as restricted to differences in the lexicon as in Borer (2014), or rejects the idea of parameters, in this sense, entirely, as in Boeckx (2010). This debate is highly complex, and so I will skip the details, but note that if a pure rationalism is to be maintained, something like parameters, accounting for linguistic variation, is necessary. It is for this reason that Boeckx advocates for a mixed approach, with an innate (rationalist) core, and variation accounted for by abstracting rules from the environment.
 31. See e.g., Sakas (2016) for a recent overview.
 32. Fodor (1975)'s position that (almost) all concepts must be innate can be understood analogously. Because there is no (known) procedure by which we can learn, i.e., *rationally* acquire, the information stored in our concepts, this information must come from within the system. Environmental stimuli can thus serve to trigger the occurrence or development of a concept, but the environment is limited to this causal role, rather than the traditional role as the source of this information.
 33. I am here focusing on internal debates about how to develop a parametric model, for which this problem is one of the most severe. Those who reject the parametric view entirely are often motivated by precisely the observations about language that I mentioned as favoring empiricist models: that languages display high degrees of variation, and these variations correlate with environmental patterns. The rejection of so-called "micro-parameters", purported parameters which correspond to these very fine-grained differences, is largely motivated in this way. See Newmeyer (2005) for discussion.
 34. E.g., A sentence which seems to be SVO can either be a result of a genuinely (underlying) SVO language (such as English), or an underlying SOV language with a rule that moves verbs into second position in surface form (such as German). There are complex empirical issues in this area. For example, if Kayne (1994) is right, then all languages are SVO in their underlying structure. This would have important repercussions for a triggering account of language acquisition, as it may make the problem of ambiguous triggers easier to solve.
 35. Other complications to the triggering model, such as the distinction between "global" and "local" triggers, i.e., triggers which unambiguously require a particular parameter setting no matter what other parameter settings are selected versus those which unambiguously call for a particular setting only *given* other settings, can be introduced in order to solve these kinds of issues. See Sakas and Fodor (2012) for a thorough proposal along these lines.

36. Although it is an empirical question whether the contribution is language-specific or not, i.e., whether language is innate in the narrow, domain-specific sense or not.
37. This fact explains something puzzling about the terminological conventions in this debate: it is the *empiricists* who are committed to the explanation of language acquisition as a *rational* activity, whereas rationalists view this process as purely causal.
38. I am skimming over significant complications in the story here. These complications should not matter for our purposes.
39. An additional complication in this debate is how we ought apportion this linguistic competence to the learner's psychological system. In particular, it may be that semi-productive rules are not acquired in the same way, i.e., by the development of the same psychological system, as the kinds of generally applicable principles and (possibly) parameters discussed in previous sections. A picture of this sort is suggested in Dupre (2019). However, I take it that an account of language acquisition in general must account for all kinds of acquisition, whether this involves the development of just one specifically linguistic system or many. More on this issue in section 8.
40. Yang (2016) Chapter 4.
41. An extra benefit is that such a proposal provides a neat explanation for historical language change. When, for whatever reason, the patterns in the environment are modified (say by the influx of speakers of different languages as a result of mass immigration), the children will pick up on, and reflect, such patterns. It is much more difficult to give a triggering-based account of this phenomenon. See Yang (2000) for discussion.
42. See, for example, Chomsky (2000) (p. 5).
43. An interesting possibility, however, would be that this mixed approach can itself resolve some of the issues with the Principles and Parameters approach. In particular, this approach seemed to flounder because inter-linguistic variation seemed to be too fine-grained, leading to the positing of too many micro-parameters, and too sensitive to the environment, as we saw in the discussion of semi-productive rules. If some of this variation and sensitivity can be viewed as acquired separately by empiricist-style models, it could be possible to revise such a picture and avoid these problems. On such a view, language variation reflects two distinct modes of language acquisition: parameter setting *and* learning. Of course, this would still fail to be a pure rationalist model. This would perhaps be a slightly messier model than those discussed in the text, but who expected cognitive science to be clean?

Acknowledgments

This paper benefited greatly from discussions with from Josh Armstrong, Sam Cumming, Guillermo Del Pinal, Bill Kowalsky, and members of the UCLA Language and Mind Workshop, as well as helpful comments from two anonymous reviewers for this journal.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Leverhulme Trust [ECF-2020-424].

Notes on contributor

Gabe Dupre is a Leverhulme Early Career Fellow, in the School of Social, Political, and Global Studies at Keele University. He works on philosophical issues in linguistic theory. Previously, he was a Teaching Fellow at Reading University, and before that he studied at the University of California, Los Angeles, and Bristol University.

ORCID

Gabe Dupre  <http://orcid.org/0000-0001-9312-4691>

References

- Ariew, A. (1996). Innateness and canalization. *Philosophy of Science*, 63, S19– S27. <https://doi.org/10.1086/289932>
- Ariew, A. (1999). Innateness is canalization: In defense of a developmental account of innateness. In Hardcastle, V. G ed. *Where biology meets psychology: Philosophical essays* MIT Press (pp. 117–138).
- Baker, M. C. (2002). *The atoms of language: The mind's hidden rules of grammar*. Basic books.
- Barkow, J. H., Cosmides, L., & Tooby, J. (1995). *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford University Press.
- Barrett, H. C. (2014). *The shape of thought: How mental adaptations evolve*. Oxford University Press.
- Bateson, P., & Mameli, M. (2007). The innate and the acquired: Useful clusters or a residual distinction from folk biology? *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 49(8), 818–831. <https://doi.org/10.1002/dev.20277>
- Boeckx, C. (2010). What principles and parameters got wrong. In C. Picallo (Ed.), *Linguistic variation in the minimalist framework* (pp. 155–178). Oxford University Press.
- Boeckx, C. (2014). *Elementary syntactic structures: Prospects of a feature-free syntax* (Vol. 144). Cambridge University Press.
- Borer, H. (2014). *Parametric syntax: Case studies in semitic and romance languages* (Vol. 13). Walter de Gruyter.
- Buller, D. J. (2006). *Adapting minds: Evolutionary psychology and the persistent quest for human nature*. MIT press.
- Burge, T. (2010). *Origins of objectivity*. Oxford University Press.
- Chater, N., Clark, A., Goldsmith, J. A., & Perfors, A. (2015). *Empiricism and language learnability*. OUP Oxford.
- Chater, N., Goodman, N., Griffiths, T. L., Kemp, C., Oaksford, M., & Tenenbaum, J. B. (2011). The imaginary fundamentalists: The unshocking truth about bayesian cognitive science. *Behavioral and Brain Sciences*, 34(4), 194–196. <https://doi.org/10.1017/S0140525X11000239>
- Chomsky, N. (1965). *Aspects of the theory of syntax*. The MIT Press.
- Chomsky, N. (2006). *Language and mind* (3rd ed.). Cambridge University Press
- Chomsky, N. (1995). *The minimalist program*. The MIT Press.
- Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge University Press.
- Chomsky, N. (2002). *On nature and language*. Cambridge University Press.

- Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36(1), 1–22. <https://doi.org/10.1162/0024389052993655>
- Chomsky, N. (2007a). Approaching up from below. In U. Sauerland & H. M. Gärtner (Eds.), *Interfaces+ recursion= language?* (Vol. 89, pp. 1–30). De Gruyter.
- Chomsky, N. (2007b). Biolinguistic explorations: Design, development, evolution. *International Journal of Philosophical Studies*, 15(1), 1–21. <https://doi.org/10.1080/09672550601143078>
- Chomsky, N. (2015). *What kind of creatures are we?* Columbia University Press.
- Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31(5), 489–509. <https://doi.org/10.1017/S0140525X08004998>
- Collins, J. (2005). Nativism: In defense of a biological understanding. *Philosophical Psychology*, 18(2), 157–177. <https://doi.org/10.1080/09515080500169686>
- Cowie, F. (1999). *What's within? Nativism reconsidered*. Oxford University Press.
- Crain, S., & Pietroski, P. (2001). Nature, nurture and universal grammar. *Linguistics and Philosophy*, 24(2), 139–186. <https://doi.org/10.1023/A:1005694100138>
- Culbertson, J., & Smolensky, P. (2012). A bayesian model of biases in artificial language learning: The case of a word-order universal. *Cognitive Science*, 36(8), 1468–1498. <https://doi.org/10.1111/j.1551-6709.2012.01264.x>
- Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford University Press.
- Curtiss, S., Fromkin, V., Krashen, S., Rigler, D., & Rigler, M. (1974). The linguistic development of genie. *Language* vol. 50, No. 3 (pp. 528–554).
- Damasio, A. R. (1994). *Descartes' error*. Random House.
- Descartes, R. (1641/2017). *Meditations on first philosophy*. Cambridge University Press.
- Devitt, M. (2006). *Ignorance of language*. Oxford University Press on Demand.
- Dupre, G. (2019). Linguistics and the explanatory economy. *Synthese* (pp. 1–43).
- Fodor, J. A. (1975). *The language of thought*. Harvard University Press.
- Fodor, J. A. (1981). The current status of the innateness controversy. In Fodor, J. *RePresentations* MIT Press (pp. 257–316).
- Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press.
- Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. Oxford University Press.
- Gibson, E., & Wexler, K. (1994). Triggers. *Linguistic Inquiry*, 25(3), 407–454. <http://www.jstor.org/stable/4178869>
- Glackin, S. N. (2011). Universal grammar and the baldwin effect: A hypothesis and some philosophical consequences. *Biology & Philosophy*, 26(2), 201–222. <https://doi.org/10.1007/s10539-010-9225-3>
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Goldin-Meadow, S. (2005). *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language*. Psychology Press.
- Goldin-Meadow, S., & Yang, C. (2017). Statistical evidence that a child can create a combinatorial linguistic system without external linguistic input: Implications for language evolution. *Neuroscience & Biobehavioral Reviews*, vol. 81 part B, 150–157. <https://doi.org/10.1016/j.neubiorev.2016.12.016>
- Gould, S. J., & Lewontin, R. C. (1979). The spandrels of san marco and the panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society London B*, 205(1161), 581–598. <https://doi.org/10.1098/rspb.1979.0086>
- Griffiths, P. (2002). What is innateness? *The Monist*, 85(1), 70–85. <https://doi.org/10.5840/monist20028518>
- Gulordava, K., Bojanowski, P., Grave, E., Linzen, T., & Baroni, M. (2018). Colorless green recurrent networks dream hierarchically. In *Proceedings of the 2018 Conference of the*

- North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 1195–1205). New Orleans, Louisiana: Association for Computational Linguistics.
- Hale, M., & Reiss, C. (2008). *The phonological enterprise*. Oxford University Press.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.
- James, R., & Blair, R. (2002). Neuro-cognitive models of acquired sociopathy and developmental psychopathy. In Glickson, J. ed. *The neurobiology of criminal behavior* (pp. 157–186). Springer.
- Katz, J. J. (1980). *Language and other abstract objects*. Rowman and Littlefield Publishers.
- Kayne, R. S. (1994). *The antisymmetry of syntax* (Vol. 25). MIT Press.
- Legate, J. A., & Yang, C. D. (2002). Empirical re-assessment of stimulus poverty arguments. *The Linguistic Review*, 18(1–2), 151–162. <https://doi.org/10.1515/tlir.19.1-2.151>
- Lewontin, R. C., Kamin, L. J., & Rose, S. P. (1984). *Not in our genes: Biology, ideology, and human nature*. Pantheon.
- Lickliter, R., & Honeycutt, H. (2003). Developmental dynamics: Toward a biologically plausible evolutionary psychology. *Psychological Bulletin*, 129(6), 819. <https://doi.org/10.1037/0033-2909.129.6.819>
- Lipton, P. (1990). Contrastive explanation. In D. Knowles (Ed.), *Explanation and its limits* (pp. 247–266). Cambridge University Press.
- Locke, J. (1836/1996). *An essay concerning human understanding*. Penguin.
- Lorenz, K. (1966). *Evolution and modification of behavior*. Methuen London.
- Mameli, M., & Bateson, P. (2006). Innateness and the sciences. *Biology and Philosophy*, 21(2), 155–188. <https://doi.org/10.1007/s10539-005-5144-0>
- Mameli, M., & Bateson, P. (2011). An evaluation of the concept of innateness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1563), 436–443. <https://doi.org/10.1098/rstb.2010.0174>
- Manning, C. D. (2003). Probabilistic syntax. In J. Hay., S. Jannedy, & R. Bod (Eds.), *Probabilistic Linguistics* (pp. 289–341). MIT press.
- Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition*, 46(1), 53–85. [https://doi.org/10.1016/0010-0277\(93\)90022-N](https://doi.org/10.1016/0010-0277(93)90022-N)
- Margolis, E., & Laurence, S. (2013). In defense of nativism. *Philosophical Studies*, 165(2), 693–718. <https://doi.org/10.1007/s11098-012-9972-x>
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.
- McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, 6(11), 465–472. [https://doi.org/10.1016/S1364-6613\(02\)01993-9](https://doi.org/10.1016/S1364-6613(02)01993-9)
- Newmeyer, F. J. (2005). *Possible and probable languages: A generative perspective on linguistic typology*. Oxford University Press.
- O'Neill, E. (2015). Relativizing innateness: Innateness as the insensitivity of the appearance of a trait with respect to specified environmental variation. *Biology & Philosophy*, 30(2), 211–225. <https://doi.org/10.1007/s10539-014-9465-8>
- Onnis, L., Waterfall, H. R., & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition*, 109(3), 423–430. <https://doi.org/10.1016/j.cognition.2008.10.004>
- Oyama, S., Griffiths, P. E., & Gray, R. D. (2003). *Cycles of contingency: Developmental systems and evolution*. MIT Press.
- Pefors, A. (2012). Bayesian models of cognition: What's built in after all? *Philosophy Compass*, 7(2), 127–138. <https://doi.org/10.1111/j.1747-9991.2011.00467.x>

- Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, 118(3), 306–338. <https://doi.org/10.1016/j.cognition.2010.11.001>
- Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, 37(3), 607–642. <https://doi.org/10.1017/S0305000910000012>
- Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, 18(1–2), 9–50. <https://doi.org/10.1515/tlir.19.1–2.9>
- Real, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science*, 29(6), 1007–1028. https://doi.org/10.1207/s15516709cog0000_28
- Rose, H., & Rose, S. (2010). *Alas poor darwin: Arguments against evolutionary psychology*. Random House.
- Sakas, W. (2016). Computational approaches to parameter setting in generative linguistics. In J. Lidz, W. Snyder, & J. Pater (Eds.), *The Oxford handbook of developmental linguistics* (pp. 696–724). Oxford University Press.
- Sakas, W. G., & Fodor, J. D. (2012). Disambiguating syntactic triggers. *Language Acquisition*, 19(2), 83–143. <https://doi.org/10.1080/10489223.2012.660553>
- Sampson, G. (1989). Language acquisition: Growth or learning? *Philosophical Papers*, 18(3), 203–240. <https://doi.org/10.1080/05568648909506320>
- Samuels, R. (2002). Nativism in cognitive science. *Mind & Language*, 17(3), 233–265. <https://doi.org/10.1111/1468-0017.00197>
- Samuels, R. (2004). Innateness in cognitive science. *Trends in Cognitive Sciences*, 8(3), 136–141. <https://doi.org/10.1016/j.tics.2004.01.010>
- Tomasello, M. (2003). *Constructing a language*. Harvard university press.
- Van Dongen, S. (2006). Prior specification in bayesian statistics: Three cautionary tales. *Journal of Theoretical Biology*, 242(1), 90–100. <https://doi.org/10.1016/j.jtbi.2006.02.002>
- Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature*, 150(3811), 563–565. <https://doi.org/10.1038/150563a0>
- Wexler, K. (2003). Lenneberg’s dream: Learning, normal language development, and specific language impairment. In J. C. Schaeffer (Ed.), *Language competence across populations Toward a definition of specific language impairment*: (pp. 11–61). Routledge.
- Wiesel, T. N., & Hubel, D. H. (1963). Effects of visual deprivation on morphology and physiology of cells in the cat’s lateral geniculate body. *Journal of Neurophysiology*, 26(6), 978–993. <https://doi.org/10.1152/jn.1963.26.6.978>
- Yang, C. (2000). Internal and external forces in language change. *Language Variation and Change*, 12(3), 231–250. <https://doi.org/10.1017/S0954394500123014>
- Yang, C. (2002). *Knowledge and learning in natural language*. Oxford University Press.
- Yang, C. (2015). Negative knowledge from positive evidence. *Language*, 91(4), 938–953. <https://doi.org/10.1353/lan.2015.0054>
- Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MIT Press.
- Yang, C. (2017). Rage against the machine: Evaluation metrics in the 21st century. *Language Acquisition*, 24(2), 100–125. <https://doi.org/10.1080/10489223.2016.1274318>
- Yang, C., Crain, S., Berwick, R. C., Chomsky, N., & Bolhuis, J. J. (2017). The growth of language: Universal grammar, experience, and principles of computation. *Neuroscience & Biobehavioral Reviews*, vol.81 part B 103–119. <https://doi.org/10.1016/j.neubiorev.2016.12.023>