Philosophical Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cphp20

Scaling up Predictive Processing to language with Construction Grammar

Christian Michel

To cite this article: Christian Michel (2022): Scaling up Predictive Processing to language with Construction Grammar, Philosophical Psychology, DOI: 10.1080/09515089.2022.2050198

To link to this article: https://doi.org/10.1080/09515089.2022.2050198

PHILOSOPHICAL PSYCHOLOGY
https://doi.org/10.1080/09515089.2022.2050198

ARTICLE

Scaling up Predictive Processing to language with Construction Grammar

Christian Michel

Department of Philosophy, University of Edinburgh, Edinburgh, UK

ABSTRACT
Predictive Processing (PP) is an increasingly influential neurocognitive-computational framework. PP research has so far focused predominantly on lower level perceptual, motor, and various psychological phenomena. But PP seems to face a "scale-up challenge": How can it be extended to conceptual thought, language, and other higher cognitive competencies? Compositionality, arguably a central feature of conceptual thought, cannot easily be accounted for in PP because it is not couched in terms of classical symbol processing. I argue, using the example of language, that there is no strong reason to think that PP cannot be scaled up to higher cognition. I suggest that the tacitly assumed common-sense conception of language as Generative Grammar ("folk linguistics") and its notion of composition leads to the scale-up concerns. Fodor's Language of Thought Hypothesis (LOTH) plays the role of a cognitive computational paradigm for folk linguistics. Therefore, we do not take LOTH as facing problems with higher cognition, at least with regard to compositionality. But PP can plausibly play the role of a cognitive-computational paradigm for an alternative conception of language, namely Construction Grammar. If Construction Grammar is a plausible alternative to folk linguistics, then PP is not in a worse position than LOTH.

ARTICLE HISTORY
Received 19 May 2021
Accepted 3 March 2022

KEYWORDS
Compositionality; Construction Grammar; Generative Grammar; higher-level cognition; language; Language of Thought Hypothesis; Predictive Processing

1. Introduction

Predictive Processing (PP) is an increasingly influential cognitive-computational framework for understanding the mind (e.g., Clark, 2013, 2016; Hohwy, 2013, 2020). PP is ambitious, as it deals with cognitive agency in general, including perception, cognition, and action. The basic idea behind this paradigm is often expressed by the slogan that "the brain is a prediction machine". PP implies a revisionary picture of cognitive agency because what we believe, perceive, etc. are "hypotheses" generated by the brain that are driven to match incoming sensory evidence. The brain is

CONTACT Christian Michel chris.michel08@gmail.com Department of Philosophy, University of Edinburgh, Dugald Stewart Building, 3 Charles Street, EH8 9AD, Edinburgh, UK
Supplemental data for this article can be accessed here.

constantly improving a hierarchically structured model based on a mechanism of prediction error minimization, which approximates Bayesian inference.

I assume for the current purposes that PP is an emerging paradigm, i.e., a set of concepts and principles that guide more specific theory and model building, not a fully-fledged theory or model.[1] As it is still emerging, no consensus exists as to its precise constituting concepts and principles. For this reason, later I need to make explicit what I take those core commitments to be.

A lot of PP oriented research has focused on perceptual and motor, as well as certain specific psychological phenomena, but PP is not yet well understood where higher cognition is concerned. Indeed, it can be seen from a recent review of the philosophically oriented literature in PP by Hohwy (2020) that PP treatments of higher cognition are still marginal. Higher cognition encompasses conceptual thought generally, and, among others, specific competencies like classification/categorization, analogy making, deduction, planning, mathematical discovery and reasoning, theory building, counterfactual reasoning, and language, as well as abilities related to social and cultural interaction, communication, and collaboration between humans.

PP theorists have pointed to the capacity of the models to which they are committed (see 2.3.) to learn and represent complex, structured world knowledge, including representations on many levels of abstraction (e.g., Clark, 2016, pp. 171–176). However, the details about those representational elements, and about the compositionality of conceptual thought and language need further development. There is, of course, some incipient work, as well as a lot of related work that is close to (and very relevant for) the PP paradigm.

As to the first type of work, for instance, Friston and Frith (2015) and Vasil et al. (2020) propose PP accounts of *communication* where agents are seen as coupled generative models. *Language* has been addressed to some extent within a PP perspective (e.g., Lupyan, 2012; Lupyan & Clark, 2015), but the treatments are limited to pointing to the value of language as a device that enhances cognition in the prediction economy, especially through linguistic labels that serve as “artificial contexts” (Lupyan & Clark, 2015) and facilitate perceptual processing.[2] However, those proposals do not spell out how exactly language and concepts are represented and processed in the mind, nor do they discuss compositionality in detail.

With regard to work on higher cognition from perspectives close to PP, Bayesian approaches have become extremely influential (see, e.g., Jones & Love, 2011; Colombo & Hartmann, 2017 for an overview and discussion). In particular, *hierarchical* Bayesian approaches (e.g., Griffiths et al., 2008; Kemp & Tenenbaum, 2009; Lake et al., 2015; Tenenbaum et al., 2011) are very relevant and have been taken on board by PP.[3] However, strictly speaking, PP is Bayesian only derivatively and approximately, and adds further commitments (on which I will elaborate in a moment). Bayesian approaches are often *computational* level accounts in the sense of Marr's (1982) three levels of description account and have the form of (acausal) mathematical equations, not neuro-mechanistic models (Colombo & Hartmann, 2017, p. 455; see also Tenenbaum et al., 2011, p. 1284; Jones & Love, 2011, p. 170). But the PP paradigm, as we will see, cuts across all of Marr's levels, i.e., it also includes algorithmic and implementation-level commitments (Sprevak, 2021a).[4]

The lack of coverage of higher cognition that explicitly deals with compositional conceptual thought and language within the PP paradigm might be a symptom of what has been called the "scale-up problem" (e.g., Silva & Ferreira, 2021):

> Furthermore, a more radical approach to cognition faces the so-called "scale-up objection", namely, the challenge of proving itself relevant for the investigation of traditional problems related to higher level cognition involving concepts such as contentful information, representational states, symbolic thought, logic, mathematics etc.

This problem seems to generally afflict all cognitive accounts that deviate from traditional symbolic computationalism, a cognitive-computational paradigm famously articulated in form of Fodor's "Language of Thought Hypothesis" (LOTH; e.g. Fodor, 1975, 2008), in which thought is syntax sensitive processing of discrete symbols.

The scale-up problem also seems pressing specifically for PP. In his influential target paper from 2013, Clark pointed out that it is still unclear how to extend the PP account to higher-level cognition (Clark, 2013, p. 201):

> Questions also remain concerning the proper scope of the basic predictive processing account itself. Can that account really illuminate reason, imagination, and action-selection in all its diversity? What do the local approximations to Bayesian reasoning look like as we depart further and further from the safe shores of basic perception and motor control? What new forms of representation are then required, and how do they behave in the context of the hierarchical predictive coding regime?

Williams (2020) recently restated the general scale-up concern of the PP community in the following way:

> As even its most enthusiastic proponents acknowledge, one of the most important challenges for predictive processing is whether the mechanisms it posits can be extended to capture and explain thought (Clark, 2016, p. 299; Hohwy, 2013, p. 3; see also Roskies & Wood, 2017).

Williams has then put forward various arguments to the effect that PP cannot account for conceptual thought (Williams, 2019, 2020). I cannot discuss and respond in detail to his objections here; Williams' nuanced argumentation deserves much more space. However, I do want to highlight that most of his objections are grounded in considerations of *compositionality* as a core feature of conceptual thought. Williams argues, among others, that PP does not have the expressive power needed for conceptual thought, because it is not "richly compositional" (i.e., as expressive as first order logic). But note that the underlying notion of compositionality is based on classical amodal symbol systems and formal logic. It is precisely the aim of the present paper to consider an alternative to this notion of compositionality.

I assume then the following motivation for the scale-up concern (for which Williams is also an example). It is generally assumed that higher cognition requires conceptual thought, which is productive, systematic, and compositional (I will expand on those notions in a moment). Mental processing is then carried out by manipulating discrete mental symbols in an algorithmic fashion. A classical computational picture of the mind fits prima facie the bill better than PP because PP is couched in terms other than discrete symbols, algorithms, or rules (see also Piccinini, 2020, pp. 125–126).

The core argument in this paper starts from the idea, for which I claim no originality, that a certain conception of language – common-sense Generative Grammar (GxG) – informs the intuitions about composition, which are then underpinned by LOTH as a cognitive-computational paradigm. By "cognitive-computational paradigm" I am referring to a set of concepts and principles that guide and constrain specific theories and models about the structure and format of mental representations and their processing mechanism. I argue then that if this common-sense conception of language is replaced by *Construction Grammar* (CxG), PP can play the role of its underpinning cognitive-computational paradigm. In other words, LOTH is to GxG what PP is to CxG.[5]

The remainder of this paper is organized as follows. Firstly, I lay out what I take PP to be committed to and compare it briefly with the LOTH paradigm (Section 2). I then recapitulate the relationship of LOTH with natural language (Section 3). In Section 4, I explain the strategy for arguing that PP can in principle meet the scale up challenge for language. In Section 5, I introduce Construction Grammar, and specifically its notion of productivity, systematicity, and compositionality. In Section 6, I suggest that PP can serve as a cognitive computational paradigm for CxG.

2. PP and LOTH as cognitive computational paradigms

2.1 What sort of a theoretical entity is PP?

As pointed out already, PP has been characterized in a variety of forms, so it is crucial to clarify what sort of thing we refer to with "PP". Such a clarification is also important because PP is sometimes criticized for being underspecified, ill-defined, impossible to verify, etc. (e.g., Litwin & Miłkowski, 2020). Those criticisms, however, presuppose that PP is a theory that can produce very specific falsifiable/verifiable predictions about target phenomena. But if we characterize PP as a *paradigm* then such criticisms miss the point. What might deserve those criticisms are, of course, specific theories of specific phenomena that make use of the core concepts and principles of PP.

I take Predictive Processing to be a cognitive computational *paradigm* that is only just emerging and still under construction. For the current purposes, I use "paradigm" in a broad sense, understood as a set of concepts and principles that provide an interpretive framework that guides and constrains the development of specific theories and models of some domain of interest. Such principles can be extracted from "exemplars of good science", of course, as Kuhn believed (Bird, 2018). Here a paradigm is not some specific empirically verifiable theory that serves as an example. This characterization of PP implies that the notion of PP is necessarily schematic.

By *cognitive computational* paradigm I am referring to such a set of concepts and principles that guide and constrain further algorithmic and implementational level accounts of the nature, format and processing of mental representations that constitute cognition. I take Fodor's Language of Thought Hypothesis to be an example of a cognitive-computational paradigm.[6] As I will use it as a foil in what follows, a brief sketch is in order.

2.2. Fodor's LOTH as a cognitive-computational paradigm

Fodor's well-known and extremely influential "Language of Thought Hypothesis" (LOTH; Fodor, 1975, 2008) is an example of a cognitive-computational *paradigm* (as opposed to a *theory*) in regard to what is being discussed here. LOTH in its deterministic version is generally considered to be a dead horse as a cognitive paradigm (e.g., Williams, 2020; Piccinini, 2020, p. 312); however, when *compositionality* is being discussed, it still serves as an influential benchmark, which captures a common-sense view, and which often operates in the form of a tacit presupposition. Also, LOTH is still very much alive in *probabilistic* versions.[7]

Aydede (1997) describes LOTH as being characterized by "meta-architectural" properties, which define a class of cognitive-computational architectures that fall under it (p. 65). LOTH, according to Fodor and Pylyshyn (1988, pp. 12–13), has the following features with regard to representational format and processing principles:

(a) representations of a system have a combinatorial syntax and semantics such that structurally complex (molecular) representations are systematically built up out of structurally simple (atomic) constituents, and the semantic content of a molecular representation is a function of the semantic content of its atomic constituents together with its syntactic/formal structure, and

(b) the operations on representations are (casually) sensitive to the syntactic/formal structure of representations defined by this combinatorial syntax.

Fodor does not provide criteria for how LOTH could be empirically verified. Rather, he motivates and then puts forward a set of concepts and principles (on a cognitive level of description) with regard to mental representations and their processing that guide and constrain the development of more specific theories and implementational models of mental phenomena, i.e., LOTH rather than a *theory*, in a strict sense, is a cognitive-computational *paradigm*.

I take the PP paradigm to be at a similar level of description to LOTH and competing with it. What is needed then is to spell out what constitutes the PP paradigm. I will highlight the fundamental differences by juxtaposing the key commitments of the two paradigms with regard to representational structure and processing principles.

2.3. The PP paradigm and its core commitments

As mentioned already, so far there is no agreed-upon articulation of the PP *paradigm*. However, from Clark (2013, 2016) and Hohwy (2013, 2020),[8] and in general from the increasing literature that makes use of the PP framework in one form or another, we can extract largely overlapping concepts and principles. Those we could consider tentatively to be the PP paradigm's core commitments.

2.3.1 Core commitments of PP

The core tenet that crystalizes from the PP literature is that the mind entertains a probabilistic, hierarchical, generative model that aims at anticipating the inflow of sensory information. The central operating principle is prediction error minimization that approximates Bayesian inference. The system adapts the model such that the prediction error is minimized on average and in the long run.

It is *probabilistic* because it represents probability distributions over "hypotheses" and inferences are carried out by approximate Bayesian inference. It is *generative* because it generates top-down predictions/hypotheses (rather than merely, for instance, classifications by bottom-up processing). And it is *hierarchical*, because the hypotheses are organized in a hierarchical structure, where higher level hypotheses are the "priors" of lower-level hypotheses. The higher levels represent regularities of larger spatial and temporal scales, i.e., more compressed and abstract information.

PP emphasizes that predictions flow top-down and are being matched by the bottom-up flow of "evidence". To be in a certain perceptual state is to issue a prediction of that state that is consistent (has a minimal prediction error) with bottom-up signals that serve as evidence. PP is a *neuro*cognitive paradigm, and its concepts and principles extend to the neural level, though, as a paradigm, those are still schematic. The PP model is neurally implemented by an interconnected hierarchy of pairs of representation and error units (consisting of a group of neurons). The prediction signals from level N are compared to the representation units on level N-1 and an error signal is generated. The error is weighted by some mechanism that estimates the reliability or relevance of the error signal. This is achieved by an estimate of the *precision* of the signal. Details do not matter here,[9] but very briefly, the precision of some variable can be determined via the magnitude of the inverse of the variance of the probability distribution of that variable. The estimated precision allows, on the one hand, for tuning down noisy and unreliable error signals; in this way the system prevents the model from being unnecessarily updated. On the other hand, incoming information that is precise and reliable should lead to adjustments of the model if the top-down prediction does not conform to it. This mechanism then serves as a tool to balance the influence of the top-down versus the bottom-up flow of information: either we rely more on prior beliefs, or we are attuned more to the sensory information.

The top-down flow of prediction signals functions as "priors" that might shape predictions on lower levels. Through this complex interplay of bottom-up and top-down information flow, the model is constantly updated on all levels based on prediction errors, which should be minimized on average and in the long run. Prediction error minimization happens all the time on all levels simultaneously. This makes processing in PP holistic because a given prediction unit is directly influenced by other prediction units in adjacent layers, and indirectly by prediction units in other layers (like a domino effect). That allows for context-sensitive processing because the state of a given prediction unit is determined by the state of many other prediction units that represent this context.[10]

As the hierarchy bottoms out at the sensorimotor level, and the focus is on the prediction of sensory input and the interaction with the environment (to get a "grip on the world", as Clark expresses it), the PP paradigm can be considered an "embodied cognition" paradigm. While embodied cognition includes many different approaches (see de Bruin et al., 2018), the common theme is the central role that the body, i.e., the sensorimotor apparatus, and its interaction with the world plays in cognition. Consequently, for the PP paradigm it is natural to adopt a modality-specific (i.e., sensorimotor) format of its representations, not amodal formats as in LOTH. At higher levels in the PP model hierarchy, those modality-specific representations are more abstract and compressed and are combined into multi-modal representations.

Table 1 summarizes the proposal for the characterization of the PP core paradigm through a juxtaposition with LOTH.

2.3.2. Possible commitments that are not part of the PP paradigm

The above characterization of the PP paradigm leaves open many aspects about the exact implementation of each of the principles. For the specification of an implementational level more detailed assumptions are necessary. For instance, how many neurons compose a prediction unit? Are the principles of PP pervasive in the brain or found only in some specific brain regions? Does PP describe the *only* type of representation and processing mechanism in the brain, or are there others? Often the hierarchical structure is constrained such that a level is only connected to the next lower and higher levels. But how central is this assumption? Could there be adaptations where connections "skip" over levels?

A further debate is related to motivating the prediction error minimization principle. Friston (e.g., 2010) relates prediction error minimization to free energy minimization. This supposedly solves the "problem of life": how can an organism evade entropic disintegration? But is this a crucial assumption – and is this link a coherent assumption at all (see Williams, manuscript)? While Hohwy seems to endorse it, Clark seems not to, at least not strongly.

Furthermore, there are a variety of proposals for an associated mathematical apparatus (e.g., Clark, 2013; Spratling, 2017) and for a specific neural architecture (e.g., Bastos et al., 2012; Kanai et al., 2015; Keller & Mrsic-Flogel, 2018; Siman-Tov et al., 2019; Weilnhammer et al., 2018). Also, many other more specific questions need to be answered to get at a falsifiable theory or model: how exactly is the prediction error minimized in the brain, by stochastic gradient descent, or other mechanisms? What is an appropriate mathematical description of the node network? What is the mechanism with which nodes are added (or deleted)[11]? Is precision weighting implemented by

Table 1. Comparison of key features of LOTH and PP as cognitive-computational paradigms.

Feature	LOTH paradigm	PP paradigm
Format of representations	• amodal • abstract • deterministic (LOT)/probabilistic ("pLOT")	• modality-specific (sensorimotor grounded) • different degrees of abstraction • probabilistic
Structure of representations	• sequential/recursive	• hierarchical network with an abstraction/ compression gradient
Processing principles	• syntax-sensitive processing/ algorithmic • local	• prediction error minimization • holistic

neurotransmitter dynamics? Which ones? And so forth. The number of open questions is daunting, which shows that PP at this stage should really be seen as a paradigm for a research programme (see also Sprevak, 2021a).

With respect to commitments to mathematical models, let me briefly refer back to Williams' objections from the introduction. According to Williams (2020), PP theorists are committed to so-called *Probabilistic Graphical Models (PGMs)*. His argument is then that those models lack the necessary expressive power for conceptual thought (very roughly: they can only represent facts, not objects and relations).[12] Details do not matter for the current purposes; the point I want to make is that the PP paradigm as I have pictured it is a *mechanistic* neurocognitive paradigm. Therefore, it does not need to commit to any unifying mathematical model at all.[13]

3. LOTH and natural language

It is worthwhile briefly revising the (abductive) core argument that supports LOTH. The purpose is to highlight a crucial point for my argument, namely how our conception of language determines our conception of compositionality.

3.1. The argument for LOTH from natural language

Simplifying very much, one important motivation for LOTH stems from the observable properties of natural language. Natural languages appear to be productive, systematic, and compositional (short: PSC) in a very explicit manner: parts (words) are assembled following certain rules into sequences (sentences). It seems that we can generate from finite means, i.e., an inventory of words and grammatical rules, an infinite number of sentences; or at least we can imagine how we could go on and on infinitely in principle (productivity). It also seems that if we can produce and comprehend sentences like "Peter kisses Mary" then we can produce and comprehend systematically related sentences like "Mary kisses Peter" (systematicity). Finally, the meaning of a sentence seems to be determined by the meanings of the words it contains and the way that they are syntactically combined (compositionality).

As language expresses thoughts, the best explanation for language having the PSC property is that thought has it as well (cf. Fodor & Pylyshyn, 1988, pp. 37–41)[14,15]

3.2. Folk generative grammar (GxG) as a presupposition of LOTH

Note that "language" in LOTH must be based on some specific *conception* of language. Then, according to LOTH, the nature of thought has the structure of language *under this conception*. Fodor's conception of language is plausibly “folk linguistics”, a common-sense Chomskian-style Generative Grammar (henceforth GxG). GxG characterizes the body of knowledge one possesses when one has the competence to speak a language. According to GxG, we hold in our memory a lexicon and (recursive) rules for combining words into sentences. This folk notion of linguistics follows directly from observing the surface form of natural language as consisting of sequences of written or spoken words (or gestures).

One plausible explanation of the origin of GxG folk linguistics is that its supporting intuitions are grounded in action (see also Dutilh Novaes, 2012). Language works with syntactic rules and words, much like an assembly line where parts are put together to form compounds. That is, we intuitively model language as discrete entities that are composed of or assembled into larger entities. This leads to a concatenative view of compositionality and makes language causally perspicuous to us. There is a sequence of physical entities, written or spoken words, for instance, and they have literally been “put together” following some rules, recipe, blueprint, or algorithm.

The important point here is that LOTH lives up to the PSC desideratum, whose force is grounded in a folk linguistic conception of language. While folk linguistics is quite perspicuous and intuitively very appealing, there are alternatives, as we will see.

4. A strategy to address the scale-up challenge for PP

The scale-up concerns have not been articulated in detail in the literature, with some exceptions like Williams (2019, 2020). But any cognitive model deviating from LOTH-based classical computational models seems to evoke a concern about compositionality. Such intuitions were also behind the well-known connectionism-symbolic computation debate. Carried over to PP, it simply is not a classical computational model that relies on the rule-based processing of discrete abstract symbols. In turn, LOTH can straightforwardly account for PSC. Hence PP needs *some* story for PSC, even if it consists of qualifying it or explaining it away.

A definitive way for PP to meet the scale-up problem for language would be to put forward a specific cognitive-computational model for the language faculty under its umbrella. Such a model/theory should be empirically supported in the strong sense that Litwin and Miłkowski (2020) are demanding (i.e., the empirical evidence should be decisive evidence for the proposed model and against contenders, and not only “compatible” with the model). Also, it should ideally be able to make novel predictions. However, my ambition in this essay necessarily needs to be more modest. I will therefore focus on sketching how PP might plausibly be a cognitive-computational paradigm for some suitable existing language paradigm. If

PP can play the role of the cognitive computational paradigm for some plausible conception of language, then PP has started to meet the scale-up challenge.

The critical point to note is that LOTH is a plausible cognitive computational paradigm only for language *understood in a certain way*. This "certain way" I have characterized as folk linguistics (GxG).

Now, interestingly, the efforts to defend different cognitive-computational paradigms, like connectionism, have often focused on showing how to *replicate* the common-sense PSC property of language. That is, for example, connectionists have often felt pressed to show how to replicate language-like thought, where they tacitly accept that language is to be understood in folk linguistic terms. In other words, many defenders and opponents of LOTH are in the grip of a specific language paradigm, folk linguistics.

The argument for LOTH from Section 3 can be teased apart into two independent claims: firstly, the normally tacit assumption that language has certain properties, those captured by folk linguistics, and secondly, the claim that the best explanation for the properties of language, whatever they are, is that thought is language-like.

Criticisms of LOTH have typically focused on undermining the second claim.[16] My strategy in what follows is different: I grant the second claim but question the first one. I suggest adopting a view on language that is different from folk linguistics. In other words, I suggest a revision of what it means to say that thought is "language-like" (and as a consequence we also get a different notion of compositionality).

Let me outline then the structure of the argument that PP does not face an in-principle scale-up problem for language based on the strategy just developed:

(I) Intuitions about a scale-up problem for PP arise because of a mismatch with the common-sense notion of composition related to folk linguistics, which follows the Generative Grammar paradigm (GxG). LOTH serves as the cognitive-computational paradigm for GxG.

(II) Construction Grammar (CxG) is a plausible language paradigm for which PP can serve as an underpinning cognitive-computational paradigm.

(III) PP can then address the challenge from productivity, systematicity and compositionality (PSC) by deference to the conception of composition of CxG.

We have already established step I) in the previous sections. Let me turn to step II (Sections 5 and 6). From I) and II) then follows III).

5. Construction Grammar and its notion of compositionality

In this section, I will briefly provide a "theoretical minimum" of Construction Grammar (CxG) for those not familiar with it. CxG is arguably the main rival of the mainstream linguistic theory, namely Generative Grammar (GxG), and differs profoundly from it. After a short general introduction, I will focus on the PSC property, which is our main concern here.

5.1. What is Construction Grammar?

Construction Grammar[17] differs from Generative Grammar in important dimensions by which we can characterize a linguistic theory: (a) the way how language is acquired, (b) what sort of knowledge linguistic knowledge is, and (c) how it is represented in the mind.

(a) According to CxG, linguistic knowledge is acquired by extracting patterns on all levels of the linguistic hierarchy (e.g. phonetic, lexical, syntactic levels) from experienced language use. What matters are learned surface structures, not inborn and hidden deep structures as in GxG. Knowledge of a language is having a large inventory of such learned patterns, which are called "constructions".
(b) Crucially, according to CxG, linguistic knowledge is not structured into autonomous modules for syntax and lexicon, where syntax is purely formal, and the lexicon contains meaningful words and expressions. Rather, CxG posits only a sort of generalized lexicon, the "construct-i-con". The construct-i-con contains not only words, but also all of the learned grammatical (phonetic, morphological, and syntactic) patterns. Grammatical patterns are considered to be not purely formal like in GxG, but also meaningful. This is a most radical deviation from a common-sense view of language. The difference between words like "cat" and grammatical patterns, like, for instance, the basic sentence form subject-predicate [S P] is the level of schematicity. Both "cat" and [S P] have a meaning. However, the meaning of the latter is, of course, much more schematic/abstract (namely something like "someone did something"), but it is a meaning after all.
(c) CxG follows an embodied cognitive paradigm. In other words, the format in which linguistic forms and meanings are mentally represented is not by amodal LOT-like symbols, but representations are *modality-specific conceptualizations*. In other words, the representations are based on and abstracted from experienced sensorimotor information. Importantly, constructions need to be understood as

> "form-
> meaning pairs". For instance, a word has a form (phonology, morphology, etc.) and a meaning (the concept denoted by that word). In the case of [S P] the form is represented, e.g., as an "experience" of the sequence of the slots with first an agent and then an action. This unified view has an economic ontology: we only need modality-specific and no amodal representations.[18] This representational ontology is important common ground with PP, as we will see.

Let me now turn to making more explicit how all those characteristics lead to a view about PSC that is different from common sense GxG.

5.2. Productivity, systematicity and compositionality in Construction Grammar

As suggested previously, the appeal of the common-sense compositionality of language understood as GxG – on which LOTH rests – most likely stems from the perspicuity of the assembly of discrete entities following certain instructions. In other words, PSC mirrors the properties arising from literally assembling atomic units into molecular wholes. Those properties are then also ascribed to the language faculty, which is metaphorically understood in this manner. Langacker expresses doubts about this conception:

> our conception of composition is greatly influenced by certain metaphors whose appropriateness for natural language cannot be accepted uncritically. (Langacker, 1987, p. 452)

In the concatenative conception of composition of GxG, the syntactic form is not supposed to contribute semantic information. The semantics of the whole is exhaustively determined by the semantics of the parts and the purely formal syntax.

CxG, as opposed to GxG, is motivated by the observation that some linguistic phenomena are best explained by positing that certain semantic properties are ascribed to syntactic structures instead of the lexicon (see, e.g., Goldberg, 1995 for the argument structure of verbs). This step dilutes the distinction between grammar and lexicon. Grammatical constructions are meaningful and linguistic entities are located on a gradient from very schematic (e.g. [S P]) to very specific (e.g. "doorknob"). CxG then paints a picture where all entities are constructions, i.e., use-based form-meaning pairs. Some constructions have schematic slots that can be filled with other constructions, which in turn might have slots that can be filled in. [S P] can be made specific by filling, for instance, the "S" slot with a more concrete instantiation, like "ANIMATED_OBJECT_NOUN" until the tree bottoms out at a specific word, like "cat". When tokening a linguistic structure, like a sentence, we get a tree-like structure – with an abstraction gradient – that bottoms out, at the level of concrete words.

CxG is characterized by *weak* compositionality:

> By recognizing the existence of contentful constructions we can save the compositionality in a weakened form: the meaning of an expression is the result of integrating the meaning of the lexical items into the meanings of constructions. (Goldberg, 1995, p. 16)

> The composite structure is an entity in its own right, usually with emergent properties not inherited or strictly predictable from the components and the correspondences between them. (Langacker, 2008, p. 164)

Strong (or *full)* compositionality, in turn, allows for predictively deriving the meaning of a composite expression from its parts and the way they are arranged. A linguistic structure is a construction if its meaning cannot be predicted from its parts or from other constructions. In the CxG picture, compositionality is graded. For instance, "jar lid" is close to fully compositional. "Laptop" cannot be understood outside the context of a metonymy (the place where the computer is typically placed stands in for the computer itself). And "understand" – composed of "under" and "stand" – is not compositional at all as we cannot predict the meaning from its parts (Langacker, 2008, pp. 169-170).

The notion of compositionality in CxG takes into account three levels of semantic contribution: the components, the construction, and, importantly, the context and rich background information:

> Virtually all linguistic expressions, when first constructed, are interpreted with reference to a richly specified situational context, and much of this context is retained as they coalesce to form established units; [...] (Langacker, 1987, p. 455)

CxG further implies a notion of *partial* productivity. In constructions you cannot fill in slots freely. It is often not predictable which inserts are allowed. For instance, consider:

(1) Mary goes to school
(2) Mary goes to work.
(3) *Mary goes to company.
(4) *Mary goes to hospital.

To the speaker it is not transparent why "Mary goes to . . . " can be combined with some but not other expressions.

Even for common-sense PSC systematicity is only partial. One can say both "Peter kisses Mary" and "Mary kisses Peter". But you can't say both "Peter reads the book" and "The book reads Peter". Some authors endorse unrestricted PSC and allow for the latter type of odd sentences, but also more radically, category mistakes like "Green dreams sleep furiously", to be meaningful and truth-value bearing – they are simply false (e.g., Magidor, 2009). But not all agree, and some prefer to rely on selectional restrictions. But the question is then how to model those restrictions. We should consider a slot in a construction not literally as an empty space, but as an abstract concept (a category) that instantiates that slot. All "allowed" instances of that slot concept can then serve as "fillers".

In sum, in CxG, common-sense compositionality is replaced by a PSC property that relies on the structure of a nested hierarchical tree network. The CxG structure and processing mode is less intuitive and perspicuous then the building-block-plus-assembly-rule account. One reason is the built-in abstraction gradient. Notice that those abstraction gradients can be straightforwardly modeled by hierarchical connectionist architectures, where information is "compressed" in successive hidden layers, very much like the operations of the visual pathway, where neurons in higher levels are sensitive to larger and larger receptive fields.

6. PP as a neurocognitive-computational paradigm for CxG

With a working understanding of both PP and Cognitive Grammar and a qualified PSC property in place, we can now complete step II) of the argument from Section 4 to the effect that PP can be seen as a cognitive computational paradigm for CxG. The aim of this section is, hence, to establish the analogy between PP and CxG with regard to representational structure and basic processing principles. This analogy underwrites the claim that PP can play the role of a cognitive computational paradigm for CxG.

PP and CxG have been developed in different research communities relying on different interests, concepts, methodologies, terminologies, and perspectives. By establishing correspondences between the two paradigms, one might run the risk of forcing one into the Procrustean bed of the other by interpreting the terms and concepts too liberally. I bite the bullet here. My ambition is not to argue that there is a formally rigorous structural similarity. Nor can I develop here in detail how CxG can be *implemented* within a PP architecture, which would be a much larger project. My ambition here is only to argue that there is a striking and *suggestive* analogy.

The following comparison will focus on the core commitments of both PP and CxG, i.e., treat them as a cognitive-computational and language paradigm, respectively, in the sense defined in Section 1.

Table 1 lists the core features of LOTH and PP. The core features of LOTH match – by design – the features of GxG, which is the reason why LOTH can serve as its cognitive-computational paradigm. I proceed to arguing that a similar analogy can be fleshed out in terms of at least six features of the structure of representations and processing principles of CxG and PP that are diametrically opposed to LOTH and GxG: 1) All linguistic representations are sensorimotor grounded. 2) The structure of linguistic representation is bipolar. 3) Representations are organized into a hierarchy with an abstraction gradient. 4) They are context sensitive. 5) Processing is both top-down and bottom-up. And 6) The knowledge of a language cannot be fully formalized.[19]

6.1. Sensorimotor grounding of conceptualizations

As explained in Section 5.1. (c), Langacker rejects the amodal view of mental representations, which has been the signature of LOTH. PP and CxG allow us to make sense of having a fully modality-specific representational system (as vindicated by neo-empiricism, e.g., Barsalou, 1999, 2009; Prinz, 2002). The "sensorimotor grounding" of representations (be it concrete concepts, abstract concepts, or grammatical structures) can be fleshed out as follows (see Michel, 2020a, 2020b). A concept in the PP view, is a certain prediction unit (at some level) conceived as a root-node plus the sub-network that depends on that root-node. The sub-network spans many lower levels in the hierarchy. The lower-level nodes represent more and more concrete features of the concept, while the root node is a most abstract, "gist"-like representation that has abstracted away from many concrete features (but retains its modal nature). The structure bottoms out at the lowest level of the hierarchy, i.e., the first layer of the sensorimotor periphery. Note that concepts can be tokened "shallowly" (cf. Barsalou et al., 2008; Simmons et al., 2008), such that they do not always reach the lowest sensorimotor level. For instance, the concept CAT is represented by a prediction unit that serves as the root node plus a tree emanating from that root node with lower-level prediction units representing many "features" of the cat (information about shape, sounds, furriness, etc.). CAT can be instantiated either gist-like (only the root node is activated), or with multi-modal imagery that is concrete to different possible degrees (the lower-level prediction units are activated, the more concrete and vivid the representation). The crucial point is that in CxG grammatical structures are also "concepts" because of their meaningfulness and are represented as prediction units.

In this sense, conceptual representations, including grammatical structures, are modality-specific representations involving sensorimotor information both in PP and CxG. The view that conceptualizations are modality-specific, extended network structures is also increasingly being endorsed in neuroscience (e.g., Hoenig et al., 2008; Kiefer & Pulvermüller, 2012; Pulvermüller, 2001).

6.2. The "bi-polar" structure of linguistic representations

As already mentioned, constructions are form-meaning pairs. I suggest that constructions correspond to pairs of associated prediction units in PP. One prediction unit represents the form, the other the meaning. For instance, the word construction [CAT/"cat"] consists of a representation of the *concept* CAT in the form of a prediction unit and a representation of the written word "cat" in the form of another prediction unit. The cognitive content of the *concept* CAT is information about cats, and the content of the *word* "cat" is information about the word form "cat", which might include its composition of letters, phonetic information, and statistical information about its statistical co-occurrence with other words, among others.

Here we get a picture in PP of two parallel hierarchical networks of prediction units, one for the form, and the other for the meaning parts of the constructions. The form hierarchy represents what we consider the "formal" linguistic knowledge, the meaning network world knowledge or knowledge that is conceptual in a traditional sense. The two hierarchical networks are laterally connected, combining the corresponding parts at all of the different levels (see also Michel, 2019; Rappe, 2021 in press). Some meaning representations might not have links to form representations (non-lexicalized concepts), and some form representations have no links to meaning representations (e.g., meaningless Jabberwocky words, nonsense sentences, or pseudo-letters).

In sum, in LOTH/GxG, to know a language is to have representations of rules (or generative principles) and a lexicon. In CxG, the knowledge of a language consists in the totality of constructions (or construct-i-con). The construct-i-con corresponds to a subpart of the total PP model, namely those prediction units that are involved in some construction, i.e., constitute form-meaning pairs.

6.3. Organization of linguistic representations in a hierarchy with an abstraction gradient

Here is a toy example of how the PP hierarchy works in principle. The prediction units at level N can be seen as abstractions/compressions over patterns of prediction units at level N-1. For instance, if level N represents a word form, then N-1 represents letters, level N-2 might represent certain edge forms (of which letters are composed) and level N-3 represents a pixel pattern (that forms edge forms).

This model can implement the construct-i-con. Take again the [[S P]/SOMETHING/ONE DOES/IS SOMETHING] construction. It can be made more concrete by replacing the S and P "slots" with more concrete expressions, e.g., [S (animate object) P(action verb)/SOMEONE DOES SOMETHING]. Still, this remains schematic as we can still make the construction more concrete, e.g., ["Peter swims"/PETER SWIMS]. This is a construct that is a maximally specific sentence that could be a possible utterance. Each slot is an abstraction over possible replacements of the slots "one level more specific". By building a tree of all possible replacements for all levels, we get a hierarchical structure with an abstraction gradient. This tree structure of more and more concrete slot replacements in CxG maps onto the hierarchical structure with an abstraction gradient of the prediction unit network in PP.

6.4. Context-sensitive processing

In CxG, conceptual representations are flexible "construals", i.e., they have a variably fine-grained structure, depending on the context of their use. Langacker says:

> One dimension of construal is the level of precision and detail at which a situation is characterized. [...] Alternate terms are granularity and resolution. A highly specific expression describes a situation in fine-grained detail, with high resolution. With expressions of lesser specificity, we are limited to coarse-grained descriptions whose low resolution reveals only gross features and global organization. (Langacker, 2008, p. 55)

A specific conceptualization consisting of the activation of some hierarchical substructure of the total network draws – in an open-ended fashion – from a set of available "domains" (the concept's "domain matrix").[20] The "domain matrix" can be seen as a pool of conceptual features that can be selected on a specific use occasion. Exactly which features are selected depends on various contextual factors (previous discourse, physical/social/cultural context, background knowledge, etc.). In sum, in CxG, a concept is a network of other concepts and the information retrieved, i.e., what other concepts are co-activated, on a given use-occasion, is context-dependent.

PP provides a computational underpinning for context-sensitive modulation of concept features. This is achieved by the precision weighting mechanism that can switch features on and off depending on their estimated reliability and relevance (Michel, 2020a). In PP we can motivate the context-sensitive modulation of concept features as a means of adjusting the representational granularity. It would not be efficient to always predict a situation with the maximum level of detail. So, both PP and CxG assign an important role to the cognitive capacity to regulate the representational granularity. While CxG merely posits such a selection, PP provides a computational sketch of how such a mechanism could be implemented.

6.5. The importance of top-down in addition to bottom-up processing

One of the main tenets of the PP paradigm is the bidirectional, top-down and bottom-up flow of information in the multilayer prediction cascade. What PP especially emphasizes is the importance and pervasiveness of top-down influences or predictions which is a feature neglected by more traditional cognitive approaches. In a striking parallel manner, Goldberg stresses the "simultaneous bottom-up and top-down processing" of constructions (e.g., Goldberg, 1995, pp. 24–25). Interestingly, she then supplies an analogy from a perceptual domain, namely vision, citing Wheeler's (1970) work, which shows that letters are recognized faster in the context of a word, i.e., the recognition (top down) of a word aids the recognition of a letter, and vice versa. This is precisely the type of example from which the PP paradigm has received significant support (e.g. Rao & Ballard, 1999).

Goldberg also discusses the *predictive role* of constructions (e.g., Goldberg, 2006, pp. 103–126). She says:

> [...] generalizing beyond a particular verb to a more abstract pattern is useful in predicting overall sentence meaning. (Goldberg, 2006, p. 105)

Take as an example the polysemous verb "get" (Goldberg, 2006, p. 106), which is a weak predictor of sentence meaning. Consider:

(a) Pat got the ball over the fence.
(b) Pat got Bob a cake.

"Get" in connection with a Verb-Object1-Object2$_{\text{path}}$ structure means a caused motion, while in connection with a Verb-Object1-Object2 pattern it signifies the transfer of something. So, there is value in representing generalizations in the form of such phrase structure constructions. Interpretations of sentences can then be supported by top-down predictions of which of the two cases we are dealing with. For instance, if we get an incomplete input like "Pat got the ball _____" we can infer that we have a caused motion construction and can predict top down that the missing word needs to be an object expressing a path.

It is fair to say that the predictive approach is not developed in much detail in CxG. But my point here is that PP would plausibly be a good cognitive-computational ally with respect to this fundamental processing principle which CxG appeals to.

6.6. CxG and PP and their formalization

One important consequence of the characteristics of CxG I have laid out is that we cannot formalize the grammar in terms of generative principles or explicit rules.[21] The non-formalizability in the form of some explicit and precise formal language is endorsed, for example, by Goldberg and Langacker[22]:

> Since language [. . .] is neither self-contained nor well-defined, a complete formal description (a "generative grammar" in the classical sense) is held to be impossible in principle. [. . .] Language does not resemble a collection of computer programs. Rather, it inheres in the dynamic processing of real neural networks, [. . .] (Langacker, 2008, p. 10)

> I have avoided using all but the most minimal formalization in my own work because I believe the necessary use of features that formalism requires misleads researchers into believing that there might be a finite list of features or that many or most of the features are valid in crosslinguistic work. (Goldberg, 2013, p. 29)

The underlying reason for not endorsing fully-fledged formalisms is that CxG emphasizes the meaning of grammatical structures, but "meaning is not easily captured by a fixed set of features" (Goldberg, 2006, p. 216).

Also, a PP model cannot be fully formalized via rules and an inventory of discrete concepts with a fixed set of interpretable features. Many prediction units are not lexicalized or do not correspond to interpretable concepts because many of them are located on levels in the hierarchy lower and higher than traditionally understood concepts. Furthermore, the flexibility and context sensitivity of the whole model is also an obstacle to a full formalization. This is, again, in opposition to LOTH/GxG, which is modeled according to a formal calculus, i.e., is paradigmatically formalizable. It is the existence of rules and the explicit manipulation of discrete symbols that makes LOTH in principle tractable. In PP however, the processing is holistic with a crucial role of top-down influences and driven by a self-organizing physical mechanism.[23,24]

Some efforts have been undertaken to computationally model CxG (e.g., Bergen & Chang, 2003; Holmqvist, 1993; Van Trijp et al., 2012). However, those do not abandon the classical computational LOTH-type paradigm in their implementational proposals. It might be more promising to endorse the PP paradigm and pursue modern machine learning methods combined with PP-specific architectures (e.g., Lotter et al., 2017; Maida & Hosseni, 2020) for a cognitive-computational implementation of CxG.

Let us take stock. All of the six features discussed in this section represent common ground between PP and CxG, while at the same time they are diametrically opposed to those of the LOTH/GxG paradigm. Therefore, it might be promising that PP and CxG join forces. PP as a cognitive-computational paradigm provides basic concepts, principles, and mechanisms that can constrain and guide the development of more specific implementational level theories and models for CxG.

7. Conclusion

Fodor's Language of Thought account (LOTH) is generally recognized as a benchmark where accounting for the productivity, systematicity and compositionality of language and conceptual thought is concerned. As Predictive Processing (PP) is not couched in terms of symbolic syntax-sensitive computation like LOTH, it seems to face a scale-up challenge regarding higher cognition.

I have argued that Predictive Processing is not in a worse position than LOTH with respect to the scale-up challenge from higher cognition if one is willing to accept a different language paradigm associated with a different notion of composition. In the same way as LOTH plays the role of a cognitive-computational paradigm for common-sense Generative Grammar, I suggest that PP can play that role for Construction Grammar. PP mirrors relevant properties of the representational structure and processing of Construction Grammar in a way that is similar to how LOTH mirrors those properties of Generative Grammar. PP can then inherit the notions of compositionality, productivity, and systematicity from CxG. The proposal is, interestingly, still a form of LOTH because it accepts that thought is language-like. The novel approach, however, is that it adopts a different language paradigm.

Notes

1. Throughout the literature, PP is characterized in many ways (e.g., Williams, 2020: theory, framework; Miller Tate, 2019: research paradigm, framework).
2. Accounts of language that share some commitments with PP are Pickering and Garrod (2013) and Pickering and Gambi (2018); however, they also do not focus on compositionality.
3. Clark, for instance, considers PP to be a "process theory" for Hierarchical Bayesian Models (Clark, 2016, p. 175).
4. In a way, PP takes seriously the concerns raised by Jones and Love (2011) with regard to Bayesian cognitive modeling, namely that it should combine with other branches of the cognitive sciences (e.g., neuroscience) and integrate mechanistic models.
5. While pairing Construction Grammar (or Cognitive Linguistics more broadly) with alternative connectionist or neurocomputational approaches is not a new idea (e.g., Feldman, 2008; Pulvermüller, 2010; Pulvermüller et al., 2013), the novel contribution of this paper is to use Predictive Processing as a paradigm alternative to LOTH.
6. *Connectionism* was broadly considered to be *the* rival paradigm to LOTH. There is extensive literature on the LOTH versus connectionism debate, especially with respect to questions around compositionality, which I cannot discuss here (see, e.g., Kiefer,

2019 for a good overview and a defense for "pure connectionism"). The debate is considered by some scholars to have reached a stalemate (see also Rescorla, 2019). Even classical computationalists and connectionists nowadays tend to move toward positions that recognize the importance of neuroanatomical and neurophysiological constraints for a full multi-level picture of cognition (Piccinini, 2020, pp. 201–202). Let me highlight that the PP paradigm, as I have characterized it here, should be seen as such a *neuro*cognitive paradigm. Cognitive theories within the PP paradigm should ultimately provide *neural mechanisms* (see Piccinini, 2020, for an extensive defense of the role of neuroscience for cognitive theories).

7. Goodman et al. (2015), Piantadosi and Jacobs (2016), and Ullman and Tenenbaum (2020) have proposed accounts of concepts and conceptual development relying on *probabilistic programs* (see also https://probmods.org), which combine structured symbolic representations and probabilistic elements. Note that that such probabilistic programming languages essentially follow the LOTH paradigm, though they add symbols representing probabilistic entities to the representational ontology.
8. Hohwy calls PP the "Prediction Error Minimization" framework.
9. The exact algorithmic and implementational level description of precision weighting is still debated (see, e.g., Sprevak, 2021b).
10. In LOTH, symbols are processed "locally", i.e., their processing is context independent. Prediction units on certain levels in the hierarchy can be seen as representing hypotheses as "beliefs" (see, e.g., Smith et al., 2021), so PP allows for context sensitive belief updating. But notice that the issue of how to relate folk psychological notions like belief, desire, intention, etc. to PP is the subject of an ongoing debate (see also Dewhurst, 2017).
11. But see, Smith et al. (2020) for a recent proposal within the PP framework. The authors propose how latent variables (which they call "concepts") in the generative model can be added or deleted.
12. Williams might indeed be right that many PP theorist commit to simple PGMs. But note that PGMs could be extended to more expressive versions, e.g., *Relational* PGMs (Getoor et al., 2001).
13. Williams considers exactly this strategy (avoiding the commitment to the PGM model) on behalf of PP but thinks that PP then loses explanatory power. However, even if this were true, it would be only for a specific *theory*, not a *paradigm*.
14. There are other arguments for LOTH (see Rescorla, 2019). However, Fodor and Pylyshyn have stressed this one in the context of the debate with connectionism. I therefore take it to be the strongest argument.
15. Note that LOTH might be a "best explanation", but only *with respect to PSC*. As Fodor himself has pointed out (e.g., Fodor, 1975, pp.197–205, 2008, Chapter 4), LOTH has shortcomings regarding *other* desiderata (which should not concern us here), so it is not the best explanation *all things considered*.
16. E.g., see, Salje (2019). Some have argued that mental "maps" can give rise to the PSC property of language (Braddon-Mitchell & Jackson, 2007). It has also been argued that connectionist structures can exhibit the PSC property in an implicit way (see Aydede, 1997), e.g., using Smolensky's (1990) tensor product representations.
17. To be more precise, Construction Grammar (CxG) is a *family* of linguistic theories (see, e.g., the overview in Croft & Cruse, 2004, Chapter 10). The different versions have in common a set of basic commitments that I denote the "CxG paradigm". I will spell out those commitments with particular reference to Langacker (e.g., Langacker, 1987, 2008) as well as Goldberg (e.g., Goldberg, 1995, 2019), as those are very elaborate and influential versions of CxG.

18. But see Michel (2020b) for a view how some modality-specific representations might appear to be amodal ones.
19. Both CxG and PP are also characterized by the probabilistic nature of their representations. However, this is not a fundamental difference compared to LOTH, given that, as already mentioned, probabilistic LOTH versions exist.
20. CxG relies here on an "encyclopaedic" understanding of meaning (e.g., Langacker, 2008, p. 38; also, Kecskes, 2013, p. 81ff), as opposed to a "dictionary" view. The "dictionary" view of meaning is roughly the classical definitional account of concepts (a set of necessary and sufficient conditions), where a concept is characterized by a (limited and fixed) set of features. The encyclopedic view holds that the meaning of a concept is potentially open-ended.
21. Of course, some formalizations might be descriptively adequate approximations for a certain range of phenomena. So, I am not claiming that formal approaches are not useful.
22. However, one version of CxG, "Unification Construction Grammar," does build on a formalization where constructions are represented by fixed sets of features. However, this approach has important disadvantages (see Goldberg, 2006, pp. 215–217 for a discussion).
23. Notice that Friston's influential Free Energy Principle (e.g., Friston, 2010) builds on a formal mathematical apparatus. However, such equations seem not a suitable level of description for language and grammar that captures the PSC property.
24. Constructions could maybe compared to species that emerge in a process that cannot be fully predicted because many contingent environmental and other factors influence the outcome.

Acknowledgments

I would like to thank Mark Sprevak, Sofiia Rappe, and two anonymous reviewers for their useful comments and suggestions on earlier drafts.

Disclosure statement

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

Notes on contributor

Christian Michel is a PhD candidate at the University of Edinburgh. His research interests are focused on issues in the Philosophy of Cognitive Science, Language, Mind and Artificial Intelligence.

ORCID

Christian Michel http://orcid.org/0000-0001-9962-5403

References

Aydede, M. (1997). Language of Thought: The Connectionist Contribution. *Minds and Machines*, *7*(1), 57–101. https://doi.org/10.1023/A:1008203301671

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1281–1289. https://doi.org/10.1098/rstb.2008.0319

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 245–283). Oxford University Press.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Bergen, B. K., & Chang, N. C. (2003) *Embodied Construction Grammar in simulation-based language understanding*. Technical Report 02-004, International Computer Science Institute.

Bird, A. (2018). Thomas Kuhn. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Winter 2018 ed.). https://plato.stanford.edu/archives/win2018/entries/thomas-kuhn

Braddon-Mitchell, D., & Jackson, F. (2007). *The philosophy of mind and cognition* (2nd) ed. Blackwell Publishing.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Colombo, M., & Hartmann, S. (2017). Bayesian cognitive science, unification, and explanation. *The British Journal for the Philosophy of Science*, *68*(2), 451–484. https://doi.org/10.1093/bjps/axv036

Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge University Press.

de Bruin, L., Newen, A., & Gallagher, S. (Eds.). (2018). *The Oxford handbook of 4E cognition*. Oxford University Press.

Dewhurst, J. (2017). Folk psychology and the Bayesian brain. In T. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing*. MIND Group.

Dutilh Novaes, C. (2012). *Formal languages in logic: A philosophical and cognitive analysis*. Cambridge University Press.

Feldman, J. (2008). *From molecule to metaphor: A neural theory of language*. MIT press.

Fodor, J. A. (1975). *The Language of Thought*. Harvard University Press.

Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. Clarendon Press ; Oxford University Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1–2), 3–71. https://doi.org/10.1016/0010-0277(88)90031-5

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews. Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. J., & Frith, C. D. (2015). Active inference, communication and hermeneutics. *Cortex*, *68*, 129–143. https://doi.org/10.1016/j.cortex.2015.03.025

Getoor, L., Friedman, N., Koller, D., & Taskar, B. (2001). Learning probabilistic models of relational structure. *ICML*, *1*, 170–177. http://ai.stanford.edu/people/nir/Papers/GFTK1.pdf

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.
Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.
Goldberg, A. E(2013). In Hoffmann, T., & Trousdale, G. (Eds.) Constructionist approaches. *The Oxford Handbook of Construction Grammar*, 1. doi:10.1093/oxfordhb/9780195396683.013.0002.
Goldberg, A. E. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton University Press.
Goodman, N. D., Tenenbaum, J. B., & Gerstenberg, T. (2015). Concepts in a probabilistic language of thought. In E. Margolis & S. Laurence (Eds.), *The conceptual mind new directions in the study of concepts* (pp. 623–653). The MIT Press.
Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). *Bayesian models of cognition*. http://repository.cmu.edu/psychology/968/
Hoenig, K., Sim, E.-J., Bochev, V., Herrnberger, B., & Kiefer, M. (2008). Conceptual flexibility in the human brain: Dynamic recruitment of semantic maps from visual, motor, and motion-related areas. *Journal of Cognitive Neuroscience*, *20*(10), 1799–1814. https://doi.org/10.1162/jocn.2008.20123
Hohwy, J. (2013). *The predictive mind*. Oxford University Press.
Hohwy, J. (2020). New directions in predictive processing. *Mind & Language*, *35*(2), 209–223. https://doi.org/10.1111/mila.12281
Holmqvist, K. (1993). *Implementing cognitive semantics*. Lund University, Department of Cognitive Science.
Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169. https://doi.org/10.1017/S0140525X10003134
Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: Predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1668), 20140169–20140169. https://doi.org/10.1098/rstb.2014.0169
Kecskes, I. (2013). *Intercultural Pragmatics*. Oxford University Press.
Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: A canonical cortical computation. *Neuron*, *100*(2), 424–435. https://doi.org/10.1016/j.neuron.2018.10.003
Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, *116*(1), 20–58. https://doi.org/10.1037/a0014282
Kiefer, A. (2019). *A defense of pure connectionism* [PhD dissertation]. CUNY.
Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, *48*(7), 805–825. https://doi.org/10.1016/j.cortex.2011.04.006
Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338. https://doi.org/10.1126/science.aab3050
Langacker, R. W. (1987). *Foundations of cognitive grammar*. Stanford University Press.
Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. Oxford Univ. Press.
Litwin, P., & Miłkowski, M. (2020). Unification by fiat: Arrested development of predictive processing. *Cognitive Science*, *44*(7). https://doi.org/10.1111/cogs.12867
Lotter, W., Gabriel Kreiman, G., & Cox, D. (2017). Deep predictive coding networks for video prediction and unsupervised leaning. *Conference paper at ICLR*. Retrieved January 5, 2021, from https://arxiv.org/pdf/1605.08104.pdf
Lupyan, G. (2012). Language augmented prediction. *Frontiers in Psychology*, *3*, 422. https://doi.org/10.3389/fpsyg.2012.00422

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, *24*(4), 279–284. https://doi.org/10.1177/0963721415570732

Magidor, O. (2009). Category mistakes are meaningful. *Linguistics and Philosophy*, *32*(6), 553–581. https://doi.org/10.1007/s10988-010-9067-0

Maida, A. S., & Hosseni, M. (2020). *Hierarchical predictive coding models in a deep-learning framework*. Retrieved January 5, 2021, from https://arxiv.org/pdf/2005.03230.pdf

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Freeman.

Michel, C. (2019). The liar paradox in the predictive mind. *Pragmatics & Cognition*, *26*(2/3), 239–266. https://doi.org/10.1075/pc.19014.mic

Michel, C. (2020a). Concept contextualism through the lens of predictive processing. *Philosophical Psychology*, *33*(4), 624–647. https://doi.org/10.1080/09515089.2020.1742878

Michel, C. (2020b). Overcoming the modal/amodal dichotomy of concepts. *Phenomenology and the Cognitive Sciences*, *20*(4), 655–677. https://doi.org/10.1007/s11097-020-09678-y

Miller Tate, A. J. M. (2019). A predictive processing theory of motivation. *Synthese*, 4493–4521. doi:10.1007/s11229-019-02354-y.

Piantadosi, S. T., & Jacobs, R. A. (2016). Four problems solved by the probabilistic language of thought. *Current Directions in Psychological Science*, *25*(1), 54–59. https://doi.org/10.1177/0963721415609581

Piccinini, G. (2020). *Neurocognitive mechanisms: Explaining biological cognition*. Oxford University Press.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002. https://doi.org/10.1037/bul0000158

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Prinz, J. J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. MIT Press.

Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, *5*(12), 517–524. https://doi.org/10.1016/S1364-6613(00)01803-9

Pulvermüller, F. (2010). Brain embodiment of syntax and grammar: Discrete combinatorial mechanisms spelt out in neuronal circuits. *Brain and Language*, *112*(3), 167–179. https://doi.org/10.1016/j.bandl.2009.08.002

Pulvermüller, F., Cappelle, B., & Shtyrov, Y. (2013). Brain basis of meaning, words, constructions, and grammar. Hoffmann, T., & Trousdale, G. Eds. . In *The Oxford handbook of construction grammar*. Oxford University Press. doi:10.1093/oxfordhb/9780195396683.013.0022.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rappe, S. (in press). Predictive minds can think: Addressing generality and surface compositionality of thought. *Synthese*. http://philsci-archive.pitt.edu/id/eprint/19848

Rescorla, M. (2019). The Language of Thought Hypothesis. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2019 ed.). https://plato.stanford.edu/archives/sum2019/entries/language-thought

Roskies, A. L., & Wood, C. C. (2017). Catching the prediction wave in brain science. *Analysis*, *77*(4), 848–857. https://doi.org/10.1093/analys/anx083

Salje, L. (2019). Talking our way to systematicity. *Philosophical Studies*, *176*(10), 2563–2588. https://doi.org/10.1007/s11098-018-1141-4

Silva, M., & Ferreira, F. (2021). Special issue on radical views on cognition: Introduction. *Synthese*, *198*(S1), 1–4. https://doi.org/10.1007/s11229-020-02959-8

Siman-Tov, T., Granot, R. Y., Shany, O., Singer, N., Hendler, T., & Gordon, C. R. (2019). Is there a prediction network? Meta-analytic evidence for a cortical-subcortical network likely subserving prediction. *Neuroscience and Biobehavioral Reviews*, *105*, 262–275. https://doi.org/10.1016/j.neubiorev.2019.08.012

Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., & Barsalou, L. W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology-Paris*, *102*(1–3), 106–119. https://doi.org/10.1016/j.jphysparis.2008.03.014

Smith, R., Ramstead, M. J., & Kiefer, A. (2021). Active inference models do not contradict folk psychology. https://doi.org/10.31234/osf.io/kr5xf

Smith, R., Schwartenbeck, P., Parr, T., & Friston, K. J. (2020). An active inference approach to modeling structure learning: Concept learning as an example case. *Frontiers in Computational Neuroscience*, *14*, 41. https://doi.org/10.3389/fncom.2020.00041

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, *46*(1–2), 159–216. https://doi.org/10.1016/0004-3702(90)90007-M

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, *112*, March , 92–97. https://doi.org/10.1016/j.bandc.2015.11.003

Sprevak, M. (2021a). *Predictive coding I: Introduction*. PhilSci-Archive. http://philsci-archive.pitt.edu/id/eprint/19365

Sprevak, M. (2021b) *Predictive coding III: Algorithm*. PhilSci-Archivehttp://philsci-archive.pitt.edu/id/eprint/19488

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Ullman, T. D., & Tenenbaum, J. B. (2020). Bayesian models of conceptual development: Learning as building models of the world. *Annual Review of Developmental Psychology*, *2*(1), 533–558. https://doi.org/10.1146/annurev-devpsych-121318-084833

Van Trijp, R., Steels, L., Beuls, K., & Wellens, P. (2012). Fluid construction grammar: The new kid on the block. *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 63–68. Association for Computational Linguistics.

Vasil, J., Badcock, P. B., Constant, A., Friston, K., & Ramstead, M. J. (2020). A world unto itself: Human communication as active inference. *Frontiers in Psychology*, *11*, 417. https://doi.org/10.3389/fpsyg.2020.00417

Weilnhammer, V. A., Stuke, H., Sterzer, P., & Schmack, K. (2018). The neural correlates of hierarchical predictions for perceptual decisions. *The Journal of Neuroscience*, *38*(21), 5008–5021. https://doi.org/10.1523/JNEUROSCI.2901-17.2018

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, *1*(1), 59–85. https://doi.org/10.1016/0010-0285(70)90005-8

Williams, D. (2019). Hierarchical minds and the perception/cognition distinction. *Inquiry*, 1–23. https://doi.org/10.1080/0020174X.2019.1610045

Williams, D. (2020). *Is the brain an organ for prediction error minimization?*. http://philsci-archive.pitt.edu/id/eprint/18047

Williams, D. (2020). Predictive coding and thought. *Synthese*, *197*(4), 1749–1775. https://doi.org/10.1007/s11229-018-1768-x