

The article below is the author submitted manuscript accepted for publication in *Synthese*.

Predictive minds can think: Addressing generality and surface compositionality of thought

Sofia Rappe

¹ *Faculty of Philosophy, Philosophy of Science and Study of Religion, Ludwig-Maximilians-Universität München, Germany*

² *Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität München, Germany*

Sofia.Rappe@campus.lmu.de

ORCID ID: 0000-0003-3343-7025

Abstract: Predictive processing framework (PP) has found wide applications in cognitive science and philosophy. It is an attractive candidate for a unified account of the mind in which perception, action, and cognition fit together in a single model. However, PP cannot claim this role if it fails to accommodate an essential part of cognition—conceptual thought. Recently, Daniel Williams (2018) argued that PP struggles to address at least two of thought’s core properties — generality and rich compositionality. In this paper, I show that neither necessarily presents a problem for PP. In particular, I argue that because we do not have access to cognitive processes but only to their conscious manifestations, compositionality may be a manifest property of thought, rather than a feature of the thinking process, and result from the interplay of thinking and language. Pace Williams, both of these capacities, constituting parts of a complex and multifarious cognitive system, may be fully based on the architectural principles of PP. Under the assumption that language presents a subsystem separate from conceptual thought, I sketch out one possible way for PP to accommodate both generality and rich compositionality.

Keywords: predictive processing, conceptual thought, compositionality, generality, unification

1. Introduction

The predictive processing framework¹ (or PP for short, see Clark 2013; Hohwy 2013; Friston 2005; Rao and Ballard 1999) successfully accounts for a wide variety of perceptual and cognitive processes in multiple domains. Vision (Hohwy, Roepstorff, and Friston 2008), body-awareness (Palmer, Paton, Kirkovski, Enticott, and Hohwy 2015), language and communication (Friston and Frith 2015; Rappe 2019), emotion (Miller and Clark 2018; Seth 2013; Velasco and Loev 2020), and psychiatric disorders² have all received explanations that appeal to the basic PP architectural principles such as hierarchical generative models, long term prediction error minimization, and precision weightings. This extension is unusual. Cognitive science so far has been characterized by specialized explanations, while predictive processing promises to unify perception, action, and cognition, fitting them into a single model (Clark 2013; Seth 2015). A unified framework could offer a single coherent system of how the mind/brain functions and a better integration of sub-fields in the cognitive sciences. As Paul Thagard notes, “the value of a unified theory of thinking goes well beyond psychology, neuroscience, and other cognitive sciences” (Thagard 2019, p. xvi). Philosophically, too, the mind has also been mostly theorized as a set of faculties, modules, or capacities all of which require their own account. The idea of a “one size fits all” framework offering a single explanatory basis for cognitive and social sciences, arts, and humanities is daring and philosophically novel. However, it remains controversial whether PP can accommodate an essential part of cognition—conceptual thought. There are other reasons to object to the unifying power of PP (see e.g. Colombo and Hartmann 2017), but conceptual thought presents one of the biggest challenges: If it cannot be explained as

¹ For an accessible introduction on predictive processing see, for example, Wiese and Metzinger (2018). For a detailed recent summary of PP literature, see Hohwy 2020, and particularly the Supplementary Table 1 where he provides a representative list of recent philosophically oriented work in PP. Clark 2015 and Hohwy 2013 present two good resources for importantly distinct detailed treatments of PP.

² For PP-based accounts of hallucinations and schizophrenia see, e.g., Adams, Stephan, Brown, Frith, and Friston 2013, Horga, Schatz, Abi-Dargham, and Peterson 2014, and Fletcher and Frith 2009. For autism spectrum disorders, see Van de Cruys, Evers, Van der Hallen, Van Eylen, Boets, de-Wit, and Wagemans 2014, Pellicano and Burr 2012, and Lawson, Rees and Friston 2014. For depression and fatigue, see Stephan, Manjaly, Mathys, Weber, Paliwal, Gard et al. 2016.

a predictive process, PP cannot pretend to provide an exhaustive account of the mind (Huang 2008; Seth 2015), certainly not for philosophers.

Several have already debated how PP can account for core characteristics of cognition such as consciousness and qualia (Hohwy 2012; Clark 2019; Clark, Friston, and Wilkinson 2019; Dołęga and Dewhurst 2020) but less is done on conceptual thinking. As Daniel Williams (2018) points out, the *standard strategy* in PP remains to treat cognitive and perceptual hypotheses as forming a common generative hierarchy, with the hypotheses related to higher-order cognitive processes situated “higher up” (*cognition-on-top* approach) or more centrally (if the hierarchy is presented as a net, rather than a ladder). As a result, the *standard strategy* mostly deflects the problem of explaining thought in PP by renouncing the cognition-perception divide (Fletcher and Frith 2009; Hohwy 2013, but see Deroy 2019).³

However, the standard strategy may be a misnomer here, as the idea that there is something special about conceptual thought remains much more standard in philosophy as well as the cognitive sciences. Williams (2018) suggests that thinking has two core properties that resist the assimilation to perception: it is *general* and *richly compositional*. Generality refers to the ability to flexibly reason about phenomena at any level of spatial and temporal scale and abstraction. Compositionality, on the other hand, refers to the ability to combine concepts into structured thoughts (ensuring that the expressive power of thought matches that of at least first-order logic). According to Williams, the architectural commitments of PP (which include an interconnected perceptual-conceptual generative hierarchy, conditional independence of its levels, and probability-based relationships between them) preclude the framework from ever fully accommodating these properties of human thought. If the standard strategy breaks down, Williams argues, the proponents of predictive accounts of the mind must either accept that PP only applies to **some** cognitive processes or propose how to explain compositionality and generality of thought outside the internal PP apparatus, by appealing to language as a public symbolic system.

³ One difference still remains: cognition requires a mechanism of decoupling from the immediate environment and is afforded by *offline simulation*, while perception is more tightly coupled with the sensory input.

But are such concessions warranted or inevitable? Could PP, with its core internal architectural commitments, accommodate the properties of generality and compositionality which Williams consider sign its limits? In this paper I propose one solution.⁴ I treat linguistic and conceptual representations as distinct (section 2) and argue that compositionality may be a *surface* property resulting from an interplay of thought mechanisms, conscious access, and linguistic machinery, rather than a property of the thinking process itself (section 3). I then sketch out a PP picture of conceptual thought that accommodates both generality (section 4) and surface compositionality (section 5). Thought and language upon my proposal are two parts of a complex and multifarious cognitive system that are **fully** supported by a PP-type architecture. The paper ends with a brief discussion of the implications of the approach and possible future directions (section 6).

2. Cutting in the thinking bundle

A necessary step is to agree first on what thought is, for it to be a challenge for PP. As noted by Williams (2018) and others (see, e.g., Fodor 2008; Harman 2015; Kahneman 2011), *thought* is an umbrella term. It applies to processes such as *reasoning, planning, deliberating, and reflecting*, which seem to have different properties, functional principles, and even goals (beyond the very general ones, such as survival of the organism). The commonality between these processes is that they all appear to be conceptual, that is, require the ability to form and manipulate concepts. With concepts, at least following a popular view, come two core properties of thought. Conceptual thought is general as “we can think and flexibly reason about phenomena at any level of spatial and temporal scale and abstraction” (Williams 2018, p. 1). It is also richly compositional, as “concepts productively combine to yield our thoughts” in a specific way (Williams 2018, p. 1).

⁴ Recently, in his doctoral dissertation Alex Kiefer (2019) proposed a more general defense of connectionist architectures against the critiques related to the productivity and systematicity challenge. Furthermore, his account relies on the distinction very similar to the distinction between functional and concatenative compositionality discussed in this paper (section 3). Both of us argue that functional compositionality is sufficient for systematicity. Yet, Kiefer’s solution lies in the properties of the representations in the vector space semantics, while I focus on the properties of the information processing in the predictive architectures.

Under such a treatment, conceptual thought resembles language: Thought is a kind of linguistic proposition, and concepts are directly associated with their linguistic labels. The process of producing thought consists then in combining linguistic labels-concepts that reside at different levels of the representational hierarchy (in PP terms). However, recent evidence from linguistics⁵ and clinical neuroscience suggests that “many aspects of thought engage distinct brain regions from, and do not depend on, language” (Fedorenko and Varley 2016, p. 132). For example, visual thinking commonly used in mathematical proofs (see e.g. Nelsen 1993), while being conceptual, does not seem to involve language (Tversky 2019). Further, some non-linguistic animals, such as cephalopods, primates, and rodents, are able to perform a range of cognitive tasks typically associated with concept-formation.⁶ Of course, the same cognitive tasks may be accomplished differently in humans and non-linguistic animals. The findings, however, do suggest that some higher-order cognitive tasks associated with conceptual thought do not, **in principle**, require capacity for language.

If being conceptual and being linguistic are not one and the same, we should avoid misattributing properties from one domain to the other. The way thought *appears* in our conscious experience, as if we were having an inner speech, may have played a role in the adoption of Language-of-Thought-like approaches to thinking (Fodor 2008). However, such experience does not correspond to the whole of our thinking activity (Heavey and Hurlburt 2008) and even less reflect the *actual mechanisms* of thinking (Machery 2005; Wilkinson and Fernyhough 2018). At

⁵ As a matter of fact, there is a long history of language theorists arguing for separation between conceptual and linguistic representations. See, for example, *Some B-theorists* in Levinson (1997, p.14).

⁶ For example, Richter and colleagues (Richter, Hochner, and Kuba 2016) have demonstrated that octopuses are able to solve simple spatial puzzles that require a combination of motor actions that cannot be understood by a simple learning rule alone, while Lauren Hvorecny and colleagues (Hvorecny, Grudowski, Blakeslee, Simmons, Roy, Brooks et al. 2007) have shown that some octopuses and cuttlefishes not just spatially represent but also conditionally discriminate. There are also studies demonstrating future-oriented tool use and at least a minimal degree of planning in octopuses (Finn, Tregenza, and Norman 2009). Even more remarkably, Fiorito and Scotto (1992) report a case of domain-general observational learning (although this finding has not yet been replicated). Similarly, rodents and primates have shown the ability to successfully perform a variety of higher-order cognitive tasks. Josep Call and Michael Tomasello (2011), for example, review a large body of experimental work on chimpanzees, concluding that they understand the goals and intentions of others while simultaneously lacking capacity for language-like representation, at least beyond trivial compositionality that does not require semantic operations (Zuberbühler 2020).

the very best, we can expect that, when it occurs, our conscious experience of inner thought aligns with the outcome of subconscious thinking processing, but even that is not guaranteed. Blindsight and visual agnosia are good examples in which one can observe the mismatch between the outcomes of *perceptual processing* and *conscious perception*. In blindsight, a person has no awareness of the stimulus but is able to act on it: they are not conscious of a mail slot but can put a letter into it. When it comes to reasoning, a similar dissociation between *decision-making process* and our *experience of it* can be found in phenomena such as choice blindness. Choice blindness shows that, under certain circumstances, people attribute decisions they have not made to themselves: when presented with the opposite of their questionnaire responses, they defend the views they said to have disagreed with. In other words, our experiences of perceptual and cognitive processes do not correspond to the processes themselves.

If thought and language, and thinking processes and their manifestations, are distinct, we need to consider where the real challenge for PP lies. If we try to explain thinking processes, the intuitions provided by our conscious, phenomenological experience of language-like thought are not good guides. On the other hand, if we want to explain this conscious experience of thought, we have to acknowledge that it may arise from an interplay of a few separate processes. Pace Williams (2018), I argue that compositionality is a manifest property and is better explained as resulting from an interplay of two distinct cognitive processes—conceptual thought and language.

3. Thinking processes do not require procedural compositionality

Starting with the hypothesis that we do not have direct access to cognitive processing, we can no longer use our conscious perception of thought to determine whether concepts indeed combine to yield thoughts (compositionality as a process) or whether it only appears that thoughts are combined from concepts in a compositional manner (*surface* or *manifest* compositionality). While PP may indeed face problems with compositionality as a *process of conceptual combination*, it may be able to explain surface compositionality as a result of interaction between thought and language, both of which may themselves be PP-based, while retaining the expressive power that seems to be necessary for thought and is typically associated with conceptual combination as a

process. The main challenge then consists in showing that compositionality may indeed plausibly be only a surface property.

The claim that one needs to argue for is that thoughts, experienced or articulated linguistically, exhibit a form of compositionality that does not necessarily perfectly correspond to the compositionality of the thinking process. But which compositionality is at stake? For the most part, natural language possesses *concatenative compositionality*. A composite exhibits concatenative compositionality if its constituents are its spatial or temporal parts. For instance, as is the case in written and spoken language respectively (García-Carpintero 1996). Thought, on the other hand, may be only *functionally* compositional, that is merely require that the composite expression has proper constituents (without imposing additional spatial or temporal requirements on them). An example of such compositionality is complex tones composed of simple sine wave components, or partials. Such tones can be uniquely separated into partials both mathematically and physically (by Fourier analysis and spectrum analyzer respectively) and explained in terms of their simple wave components, despite not being temporal or spatial parts of the complex tone (García-Carpintero 1996). Importantly, García-Carpintero (1996) argues, systematicity and productivity of thought can be supported by functional compositionality alone as the ability to form relational structures or concepts does not depend on the sequential order of information processing. As Kiefer (2019, p. 234) notes, “there is no *a priori* requirement that in order to represent objects and properties, a representational system must possess separable syntactic constituents that have them as their semantic values”. For example, the system may represent relations not as individual nodes directly corresponding to the linguistic labels that we use to describe these relations, but instead as a variety of the corresponding states of affairs. Such states, however, may be further collectively generalized to yield the relevant abstract concepts and associated labels for the purpose of linguistic expression (more on the nature of concepts in PP in section 4.3 and on the interaction between thought and language in section 5.1). Although this type of representation may be rather “unwieldy and redundant”, it is not in principle impossible (Kiefer 2019, p. 234).

Because our conscious experience of thought is often mediated by natural language (Frankish 2018), and natural language syntax/semantics are, for the most part, concatenatively compositional, thought inherits the appearance of concatenative compositionality from language.

However, concatenative compositionality of language is a property of conscious structured thoughts and does not necessarily reflect how thought and language are processed in the brain. It is tempting and intuitive to link concatenative compositionality to semantic *atomism*, or the idea of progressive bottom-up conceptual or linguistic combination as a way of constructing higher order chunks of language or meaning. However, as noted by many philosophers of language, including Gottlob Frege (1879), semantic atomism quickly runs into problems. To begin with, certain linguistic concepts do not have meaning in isolation, and hence, trying to interpret them bottom up is pointless. Further, the ability to substitute parts of a sentence, which is often thought of as equivalent with bottom-up compositionality (Hodge 2001; Janssen 2001), does not hold generally (Frege 1892). One notable exception to the substitution property, for example, is subordinate clauses, where an expression can only be substituted with an expression with the same customary sense—the way the expression **presents** the referent (as opposed to its truth value). Semantic atomism does not sit too well with novel metaphors either. In metaphors, the meaning of an expression is not equal to the compositional combination of lower-level chunks or concepts, so it is hard to explain how such meaning arises through bottom-up conceptual combination (at least without significantly complicating the story).

An alternative approach, *top-down contextuality*, suggests that rather than starting with concepts and putting them together to form coherent thoughts and judgements, one obtains meanings of the parts of an expression by *decomposing* the thought (Hermes, Kambartel, Kaulbach, Long, and White 1981). The point of departure is a complete thought, and the idea is understood before individual words are recognized (Janssen 2001). This form of *top-down contextuality* may be too strong as parts of the sentence obviously contribute to the expression of the sense of the sentence (Gabriel, Hermes, Kambartel, Thiel, and Veraart 1976). However, the word “contribution” suggests that the meaning of a compound sentence is, perhaps, gestalt-like (more than the sum of meaning of its parts), as opposed to purely compositional (all elements of the meaning are contained in the lexical items and their syntactic relations within the sentence).

This weak version of compositionality is gaining popularity due to the new findings in cognitive linguistics and recent focus of semanticists on figurative speech and non-literal meanings. PP, with its simultaneous bottom-up and top-down processing style, is in a great position to accommodate these insights, especially under the idea of separation of conceptual thought and

language. In fact, as pointed out by one of the reviewers of this manuscript, the rejection of semantic atomism is explicitly argued for in defense of PP-based semantics by Kiefer and Hohwy (2018) and is generally assumed in widely popular vector space semantic models (Kiefer 2019). Before outlining my approach to compositionality in PP, however, let us tackle the generality challenge.

4. The generality problem

4.1 Personal level beliefs can still be about small objects and rapidly changing regularities

Generality targets “the fact that we can think and reason about phenomena at *any* level of spatial and temporal scale and abstraction” (Williams 2018, p. 2) in a way that flexibly combines representations across such levels. Williams argues that the existing predictive views about cognitive representation cannot accommodate this fact because of the two core commitments of PP. First, representational hierarchy tracks computational distance from sensory surfaces (higher levels predict and receive error signal from the lower levels). Second, it tracks representations of phenomena at increasingly larger spatiotemporal scales (Hohwy 2013). Following Jona Vance (2015), Williams argues that these two commitments are in tension with the standard strategy that assumes common hierarchy of perception and cognition:

[If] beliefs are supposed to exist higher up the hierarchy and moving higher up the hierarchy is supposed to result in representations of phenomena at larger spatiotemporal scales, it should be impossible to have beliefs about extremely small phenomena implicated in fast-moving regularities [...] This point doesn’t apply to uniquely ‘high-level’ reasoning of the sort found in deliberate intellectual or scientific enquiry: patients suffering from ‘delusional parasitosis’ wrongly *believe* themselves to be infested with *tiny* parasites, insects, or bugs—a fact difficult to square with Fletcher and Frith’s (2009) suggestion ... that delusions arise in ‘higher’ levels of the hierarchy, if this hierarchy is understood in terms of increasing spatiotemporal scale (Williams 2018, p. 16).

The term *beliefs* discussed above refers primarily to a folk-psychological notion of beliefs typically associated with the ‘personal level’. Such beliefs, according to the standard strategy, are

indeed thought to be located at the higher levels of the generative hierarchy (or hierarchies). However, as Williams himself notes, “there is a sense in which all information processing within predictive brains implicates *Beliefs*...As such, *the same fundamental kind of representation* underlies representation in both sensory cortices and cortical regions responsible for intuitively ‘higher-level’ cognition”. (Williams 2018, p.12). The difference between such higher-level beliefs and lower-level percepts may lie, for example, in conscious or metacognitive access, complexity, and resulting phenomenology, but not necessarily in the way of representing (although such possibility is not entirely out of the question) (Barsalou and Prinz 1997; Goldstone and Barsalou 1998; Tacca 2011).

The two commitments regarding representations above (tracking computational distance from the sensory surfaces and increasing on the spatio-temporal scale) are indeed often considered to be part-and-parcel of the standard strategy in PP. However, there is no real tension between them and the standard strategy. Rather, the tension pointed out by Vance (2015) stems not from the representational commitments themselves, but from (mis)treating the properties of representations (computational distance and spatiotemporal scale) as directly reflected in the updating process. First, let us consider computational distance. The bottom representational levels of the hierarchy are linked to the sensory inputs. As we move further along the hierarchy and further away from the sensors, we get progressively more complex representations that rely on multiple types of lower-level features. For example, lower levels of the visual hierarchy may represent simple features such as luminosity, edges, or shapes, while higher levels—multi-feature objects or an entire perceptual scene. At a certain point, representations may become multisensory.⁷ As higher-level representations are informed by prediction errors coming from the levels below, complexity, in a certain sense, implies more relevant variables. In a bottom-up processing approach this would also imply a more involved computation. However, when it comes to processing, in PP representations are not constructed and unpacked bottom-up by necessarily settling the lower levels of the hierarchy first. Instead, there is often significant top-down influence. Representations are taken to be conditionally independent from the levels not directly above or below them and updating across all levels is simultaneous, not sequential.

⁷ Although, in a certain sense, **any level** may be informed by top-down predictions (which in turn may be informed by the sensory information from other sensory domains).

Further, prediction error propagating upwards from any level may have a system-wide effect. Of course, certainty or precision of the higher-level hypotheses to a significant extent depends on accumulating prediction error and feedback from the levels below and hence, in some cases, higher levels may take longer to *settle*, but this is situation-specific and not true in general. In other words, at least in principle, the content of the higher levels can be updated as flexibly and dynamically as that of the lower levels despite the difference in computational distance to the sensory surfaces.

Let us now turn to the spatiotemporal properties: Can PP accommodate higher-level representations about lower-level representations given that representations increase in the spatiotemporal scale as we move higher up the generative hierarchy? The specific example (having delusions about being infested with tiny parasites) chosen by Williams to present this challenge is unfortunate, as it ties the notion of spatial increase in the hierarchy to the physical dimensions of the represented objects. Principally, there is no difference for PP whether one has delusions about tiny parasites or giant dinosaurs. What seems like a more relevant challenge is the span (time property) and scope (space property) of the hypotheses' content. To illustrate the point though the temporal case: If higher-level representations are in some way temporally spread-out, how can they represent temporally fine-grained (fleeting) things or properties as they exist in real time? The key is that lower-level representations (say, object-level) are represented at the higher levels as part of a more complex representation that places them within a scene/world⁸. Here, as in the case of the computational distance commitment, the way a PP-system processes information becomes relevant. Because the system is updated simultaneously across the hierarchy and does so in a largely conditionally independent manner, the object-in-a-scene representations are updated as quickly as the object representations, as long as the relevant information is available. Hence, ultimately, although lower levels are, of course, relevant to formation of the higher-level representations, neither the specific computational distance from sensory surfaces nor the position on the spatiotemporal scale do not play much role in PP's ability to support high-level hypotheses about low-level dynamically changing events.

4.2 Concepts are meaningfully located at a specific region of a hierarchy.

⁸ As kindly pointed out by one of the anonymous reviewers (verbatim).

Could we assume that *multimodal* or *amodal* representations can effectively predict the phenomena represented across different sensory modalities? This could happen, for example, by implementing *spherical* or *web-like architectures* (with proximal inputs across different modalities perturbing the outer edges, Penny 2012; but see also Gilead, Trope, and Liberman 2020), Still, according to Williams, this solution faces additional problems:

Either [PP] simply abandons the view that conceptual representation is meaningfully located at some region of a hierarchy in favour of a conventional understanding of concepts as an autonomous domain of amodal representations capable in principle of ranging over *any* phenomena [...]; *or* it offers some principled means of characterising the region of the hierarchy involved in non-perceptual domains. (Williams, 2018 p.16

These alternatives seemingly leave the standard strategy at an impasse. According to Williams, to abandon the idea of meaningfully localizing conceptual representations in the hierarchy is to contradict the standard strategy, as we can no longer talk about cognitive representations being located above the perceptual ones. On the other hand, one may not be able to find adequate criteria for localization of conceptual representations. “Do my thoughts about electrons activate representations at a different position in “the hierarchy” to my thoughts about the English football team’s defensive strategy, or the hypothesised block universe? If so, by what principle?” (Williams 2018, p. 17). I argue that it is senseful to talk about conceptual representations or (winning) hypotheses associated with personal level beliefs, thought, etc., as being located further away from the sensory cortices than non-conceptual perceptual representations without necessarily having to characterize the specific “region of the hierarchy *involved* in non-perceptual domains” (Williams 2018, p. 16). In fact, attempting to specify such a region in a certain sense would mean to fundamentally misunderstand the core tenets of PP.

To begin with, although there may be many situations where the lower perceptual levels do not play much role in cognitive inference (e.g., offline simulation), in principle, any level of the generative hierarchy (assuming it has enough weight) can affect processing at any other level. On the other hand, such approach does not either preclude or necessitate activation of sensory cortices when it comes to thinking about concepts related to perception. The standard strategy merely postulates that higher-order cognitive beliefs are located above perceptual ones, not that there is a certain restricted region that is exclusively involved in non-perceptual domains or that

perceptual information must necessarily be used to arrive at a higher-level (cognitive) hypothesis. Williams' example (2018, p. 16) of a blind person thinking about light patches is meant to present a challenge to the standard strategy but it fails as PP does not require activation of the relevant part of the visual cortex in order to conceptualize about light patches. Depending on each individual case (innate or acquired blindness, etc.) the generative model and the processing path may differ. This, to a certain extent, is true even for the people without visual (or other sensory) impairments. Generative models are specific to each individual and their learning paths, and even within one individual hypotheses about similar events may be formed in different ways case by case, with the error signal coming from any level of the hierarchy.

Another complication in characterizing a specific region involved in non-perceptual domain or allocating concepts at certain levels in absolute terms is that lower-level hypotheses contributing to conceptual representations have complex branching structures with branches of different depth and potentially incorporate information from multiple sensory domains. For that reason, it would not be possible to pinpoint a specific level “n” across the hierarchy that would purely consist of representations of concepts (or any kind of representations of certain complexity or united by certain features). None of this, however, precludes us from making sense of the cognition-on-top approach in PP. Rather than fixing the location, we just need to specify how concepts arise and relate to the contributing lower-level perceptual hypotheses.

4.3 Concepts are dynamic representations

So, what would be a helpful way to think about concepts in PP with respect to the considerations above? Following Michel (2016, 2020), I suggest that concepts in PP may be best thought of as “highly flexible, dynamic, and context-dependent representations” (Michel 2020, p. 625), rather than relatively stable theories, and include ad-hoc, non-consciously accessible, multi-modal representations with cross-domain connections (Michel 2016, 2020). A concept of a cat, for example, can be linked to various features such as fluffy texture, purring sound, and triangular ears, as well as a range of “cat-related” situations. Conceptual representations, however, are “thinner” than the contributing representations and do not contain all the richness of detail. According to Michel (2020), the cognitive role of concepts is precisely in abstracting away the unnecessary, irrelevant information. This allows the brain to more efficiently generate error-

minimized predictions given the high cost of metabolic activity. For example, in order to avoid stepping on a cat, it would be sufficient to have information about its rough shape, size, and some general patterns of behavior (Michel 2020). In some cases, entirely new prediction units may be created on the fly for abstracting a frugal representation (Michel 2020, Kiefer 2019). To qualify as concepts in the traditional sense of the word, such representations would need to additionally gain conscious accessibility (perhaps, through language), relative stability, and applicability across a certain range of domains.

A thin and fleeting concept could grow into a rich and stable one if it turns out to be useful in the prediction economy. Also, representations that initially have a narrow range of application might get a more generalized use through mechanisms like “neural recycling.” (Michel 2020, p. 635)

The existing concepts can continually fine-tune their cognitive content and modulate the content used for inferences depending on the context. Picking the relevant parts of the cognitive content in each context and selecting the most efficient route for categorization may be afforded by precision-weighting. To summarize, concepts are predictive units crucial for “data compression and context-sensitive modulation of the prediction detail” (Michel 2020, p. 634). To that, I would add that such compression is not only beneficial for efficient processing, but also for communication. As a speculative point, this may be the reason why language operates with lexical units that tend to correspond to stabilized concepts. This approach to concepts as dynamic multi-modal representations (although not in the PP context) is further supported by recent finding on distributed concepts (Handjaras, Ricciardi, Leo, Lenci, Cecchetti, Cosottini et al. 2016), re-wiring experiments (Newton and Sur, 2005), and neural re-usage phenomena (Anderson 2010) that suggest a high degree of flexibility and interconnectivity (Michel 2020).

Importantly, such an approach does not contradict either the idea of spatiotemporal scale increase when it is understood as hypothesis generality, or Williams’s suggestion that abstract concepts are not strictly causal (they are aggregations constructed for efficiency, but not necessarily something meaningfully located in the world). Further, the notion of concepts as dynamically generated hypotheses abstracting features of the lower-level representations still allows concepts to be meaningfully located in a region of the hierarchy above hypotheses directly related to

sensory input. Concepts are essentially multimodal, cross-domain representations constructed on top of the non-conceptual perceptual ones, and the more abstract ones are further removed from their perceptual beginnings (see, e.g., Meteyard et al. 2012, Kiefer and Hohwy 2018). The dynamic “made to order” representations, on one hand, extrapolate from perceptual features and may serve as a proper top level of the perpetual hierarchy, essentially forming our perceptual ontology. On the other hand, they serve as constituents of thought and stabilize into less dynamic “full” concepts. This is in line with the weak embodiment approach to concepts (see, for example, Meteyard et al. 2012 and Dove 2018), works nicely with the traditional “cognition-on-top” PP, which seems to presuppose some kind of transitional area (levels) that would bridge perception and cognition, and accommodates the observation that perceptual categories are much more context dependent and flexible than concepts **proper** (Deroy, 2019). That said, where specifically the border between conceptual and non-conceptual representations lies in perception is a topic of an ongoing debate. One extreme view is to argue that all perceptual representations in the hierarchy with the only exception of the lowest level (directly representing the incoming sensory signal) could be called conceptual. In such case, conceptual representations would span almost the entirety of the generative hierarchy. Yet, it seems that for most people, including Michel (2020), when it comes to conceptual representations at least a certain degree of abstraction is implied, which means that concepts are located, for example, above the specific structured collections of visual features that we recognize (with the help of concepts or not) as instances of specific objects. This interpretation is consistent with the standard strategy in PP and puts concepts on top of the non-conceptual perceptual levels.

But how does this picture result in the kind of seemingly compositional expressions combining “representations from across levels of any conceivable hierarchy” that we register as thoughts (Williams 2018, p. 17) on the conscious level? This question brings us back to the discussions of compositionality and the interaction between language and thought.

5. Separation between thought and language in PP may address the compositionality problem

Let us return to William's notion of *rich compositionality*. According to Williams, compositionality is a "principle of representational systems in which the representational properties of a set of atomic representations *compose to yield* the representational properties of molecular representations" (my emphasis, Williams 2018, p. 18). Compositionality of the representational system is further commonly taken to explain productivity and systematicity of human cognition, that is our capacity to come up with an infinite number of meaningful utterances and the fact that the ability to produce some thoughts is inherently tied to the ability to produce others. Nevertheless, as Williams notes, it is not enough to express compositionality in terms of productivity and systematicity as all sorts of cognitive architectures that are otherwise limited can potentially produce them in different varieties (Williams 2018). For example, both productivity (infinitary character) and systematicity underlying thought only require functional systematicity (see section 3). As discussed in section 2, due to the lack of access to, or representative capture of the underlying mechanisms, even the assessment of directionality of the process cannot be safely made. Whether in thought atomic representations compose to *yield* the representational properties of molecular representations is contentious.

To more strictly specify the compositional requirements for thought, Williams proposes a definition of rich compositionality, that is the property of being *at least as expressive as first-order logic*. (Williams 2018). As he notes, this seems a fairly minimal requirement: "Notice how low the bar is in the current context: all one requires is to show that *some* features of higher cognition are at least as expressive as first-order logic. This claim could not plausibly be denied" (Williams 2018, p. 21). He further argues that PP is unable to satisfy this requirement due to its commitment to the kind of connectionist architecture that may be represented through hierarchical probabilistic graphical models. Such graphical models have expressive power equivalent to propositional logic, that is limited to facts with "operations [...] defined over atomic (i.e. unstructured) representations of such facts—namely, propositions" (Williams 2018, p. 20). The ontology of first-order logic, on the other hand, comprises not just facts but *objects* and *relations*, thereby representing "the world as having *things* in it that are *related* to each other, not just variables with values" (Russell and Norvig 2010, p. 58, as cited by Williams 2018).

Williams’s proposed strategies are either to abandon the specific predictive coding architecture, restrict the scope of the framework, or to relegate the explanation of the aforementioned phenomena to the “distinctive contribution of natural language and *public* combinatorial symbol systems more generally” (Williams 2018, p.22). The challenge with the latter strategy is to explain how “human thought... inherit[s] such systematicity as it displays from the grammatical structure of human language itself” (Clark 2000, as cited by Williams 2018, p. 23). In light of the discussion in section 3, I would like to reformulate this challenge not in terms of inheritance or transfer, but in terms of *interaction* between thought and language. The lateral connections and non-homogeneous priors so strongly emphasized by Clark allow for a variety of multidimensional subsystems (including linguistic) that integrate with each other at various points and whose existence, nevertheless, does not contradict the commitment to the predictive coding architecture. In the following paragraphs, I outline a PP-type picture that explains how conceptual and linguistic representations work together to produce the kind of introspective experience of thought that we have.

5.1 PP-based proposal of how conceptual thought and language interact

Consider a PP hierarchy of inference (*perceptuo-conceptual hierarchy* or PCH) in line with the standard strategy. Conceptual thought in the form of hypotheses is distributed across different levels. Each node below the top one represents a lower-level hypothesis and is taken to be causally dependent on the adjacent levels only. The processing system operates simultaneously both in top-down and bottom-up fashion progressively updating the precision-weighted predictions supplied in the downward flow. As discussed in the previous section, concepts in PP are representational systems that form dynamically as hypotheses and include ad-hoc, non-consciously accessible, multi-modal representations with cross-domain connections that help to efficiently generate error-minimized predictions (Michel 2020). Concepts representing relations between the fact-type hypotheses (described by Williams as the only kinds of hypothesis available to PP) are similarly just hypotheses/concepts statistically extrapolated from individual instances involving such relations. Rich dimensionality of the generative models (and human neural networks) allows for such dependencies to be integrated as separate inferential nodes.

Certain nodes in the PCH (especially at the level of stabilized concepts) may be associated with linguistic tags on a probabilistic basis. Such matching can happen either simultaneously with the

inferential process in the PCH or after the inference stabilizes—recall the idea of decomposition of thought to obtain meaning of separate parts (section 3). The former seems to me like the more likely option, although this question should be settled empirically.

The rules and regularities of language use may form their own part of the hierarchical generative model (LH). The representations in the LH are essentially of the same type as in the PCH. However, the interaction between LH and PCH parts of the model is characterized primarily by lateral connectivity, while, internally, LH and PCH are characterized by a more traditional, primarily hierarchical architecture. The lateral connections between conceptual and linguistic hierarchies occur at various levels (starting with the level where relationship between concepts and words may be identified). However, not all concepts (whether of relationship or object type) have to directly map onto linguistic tags used to express thought. It is sufficient that the conceptual system *overall* can be mapped to a corresponding linguistic system. Successful association of these objects/concepts with the linguistic tags allows the system to start the sentence generation process, which is done in a PP simultaneous bottom-up and top-down manner.

In fact, element-wise one-to-one mapping between LH and PCH may be impossible because thought and language may have different consistency and coherence requirements. Concepts, for example, may be largely redundant and reemerge in different locations in the PCH.⁹ In fact, the picture of the PCH introduced above makes it unlikely that conceptual representations would form “a coherent and consistent body of knowledge that we can fully formalize in a propositional or language-like format” (Michel 2020, p.636). As Michel points out, some argue that such inconsistency is not only a necessary property of a human mind (Sorensen 2004), but a desirable evolutionary feature in a highly uncertain environment (Bortolotti and Sullivan-Bissett 2017). Natural language, on the other hand, requires (with some exceptions) consistency and coherence. Perhaps, “we should view formal systems like [languages] as cultural artifacts that contribute to

⁹ As suggested by one of the anonymous reviewers, this redundancy may also point to a solution to the generality challenge. For example, the rapidly changing edges in the visual field may be represented directly in early vision close to the sensory periphery, but also higher up in the hierarchy as a proper concept. This would help to differentiate between the level of abstraction of the regularity vs. representation. The reviewer’s concern is that it may be rather ad hoc to supplement PP with such additional distinctions, but see the discussion directly following this footnote in the main text.

shaping the mind rather than constitute it” (Michel 2020, p.636; see also Dutilh Novaes, 2012, p. 61). Now, to illustrate the model above on a concrete example, consider the following sentence: A cat is on the mat (figure 1).

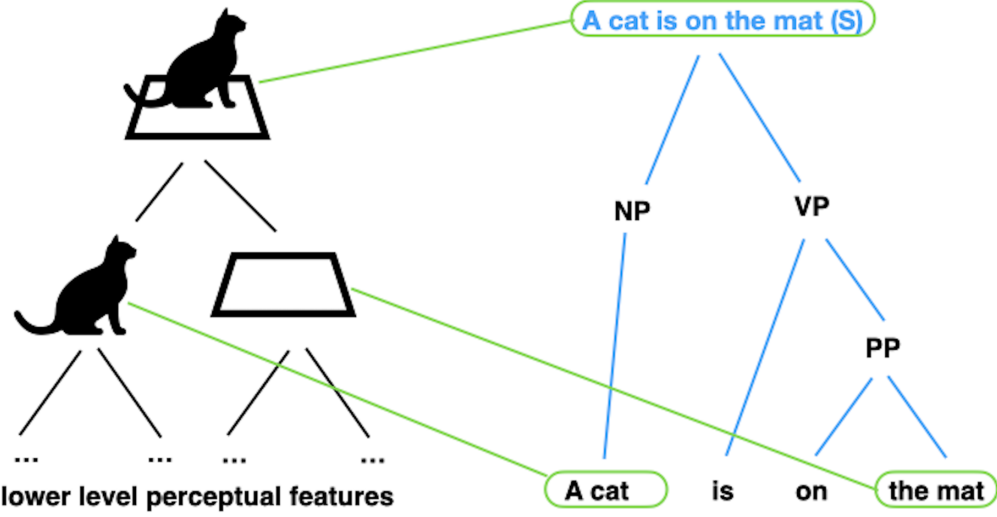


Fig. 1 A cat is on the mat

This may be either a description of the ongoing perceptual experience, or a thought that is isolated from the current sensory input. The former is the standard case of perceptual processing. In order to generate a sentence that describes the scene (perhaps, describes best given a specific context) we require nodes in the perceptual hierarchy that can ‘anchor’ linguistic tags. An obvious contender here are object-level hypotheses (cat, mat) presenting instances of dynamic representational concepts of the type described above. Prepositions, such as ‘on’, may also have statistically associated lateral links to the relational concepts in the PCH. The mapping, however, does not require for the relational concept to be as straightforward as the proposition itself—the latter is associated with the former on a good-enough basis. Further, the associated relational concept may be positioned above object concepts in the PCH. The mutual relationship between concepts in the hierarchy does not matter as the linguistic system is connected to the PCH laterally (green lines in figure 1). Other lexical items like, ‘is’, ‘a’, or ‘the’ may be potentially treated similarly (as they may be taken to be meaning-bearing), but in general, as discussed, not all the words have to correspond to some nodes in the PCH — Grammar requirements may be settled within and expressed locally in the predictions running through the LH (see Rappe 2019).

For example, the ‘cat’ node in LH is associated with the cat representation in the PCH, but as part of LH is also associated with the specific rules of use. These rules of use are statistical properties — strong predictions about grammatical context in which these nodes may occur. Explicit rules may be abstracted from these regularities, but the idea here is that we originally learn “language in use”, and the explicit rules follow later (if they are at all explicitly represented).¹⁰ Observation of the semantic-syntactic rules of natural language can lead to the introduction of new sentence parts that do not have direct correspondents in the PCH.

Linguistic expression stabilizes both based on the LH feedback loop and the prediction error coming from the PCH. Linguistic hierarchy has a significantly more rigid structure and captures thought as a kind of net stretched over the generative model. For the linguistic expression to match the content of the winning inferential hypothesis, linguistic subhierarchy must arrive at the state with undetectable relevant prediction error internally (which ensures coherence of the linguistic utterance) but the lateral communication between LH and PCH must also stabilize.

This relationship between PCH and LH is mutual with both hierarchies able to influence each other. The situation is very similar for the case of thought that does not represent the current sensory input, except for the hyperprior that the lower-level sensory information does not play as important a role for error generation. The mutual bootstrapping relationship between PCH and LH in this case also helps address Williams’s point that a concept cannot be generated without the activation of the sensory path. That said, generation of imagery and phenomenological effects such as tasting, smelling, and experiencing tactile sensations may often accompany thought. Perhaps, this could be explained by the downstream effect of prediction generation.

5.2 Surface compositionality and implications of the approach

Inferential hierarchies are functionally compositional, with top-down prediction propagation. The appearance of bottom-up and concatenative compositionality (at the manifestation level) in thought is largely due to the fact that thoughts, when they manifest at the conscious level, typically take the shape of natural language. When we receive language input, it unfolds in a

¹⁰ Although figure 1 represents linguistic hierarchy in the old-fashioned syntactic-tree style, the approach does not presuppose any specific way of encoding linguistic rules. The representational model above is chosen simply for convenience.

temporally (oral) or spatially (written) spread-out manner. Due to the jigsaw-puzzle-like nature of grammar (allowing for composition of higher-order units from the lower-level ones by directly combining them), thought appears to have concatenative compositionality. This is also true for language production as it requires uttering linguistic units one by one. Both inferential models and linguistic processing models are updated in the bidirectional fashion as prescribed by PP architecture.¹¹ However, this is done in a sort of parallel-processing fashion with concept-to/from-linguistic-unit bridges often serving as the lowest-level connections.

Importantly, both language and thought influence each other. Linguistic tags help to improve certainty of the ongoing inference, but they can also evoke or activate inference through their relations to concept-hypotheses. The inference-language loop may also function as an additional source of prediction errors to match the hypothesis against during offline simulation and serve as a kind of bootstrapping mechanism.

The picture described above captures productivity and systematicity of thought as well as functional compositionality in inference and concatenative compositionality in language. It also allows us to explain why concatenative compositionality does not hold in some cases, such as that of metaphor: language is about best approximation of thought **overall**, but it is not about building thought from the ground up. Although both thought and natural language hierarchies ultimately describe the same states of affairs, the mapping is not necessarily one-to-one.

To summarize, interconnectedness of thought and language is what creates the illusion of concatenative bottom-up compositionality in conceptual thinking. The requirement of separate conceptual and linguistic hierarchies may resemble Williams's second strategy where the properties of compositionality and generality are delegated to language and public symbolic systems. The crucial difference is that all parts of my sketch are purely PP-based. For this reason, despite introducing multiple subsystems, we may still be able to talk about a single unified model implemented by the brain. This, however, requires one to accept unification at the computational (as opposed to the implementational) level. After all, both conceptual and linguistic hierarchies

¹¹ For more on language processing within the PP paradigm, see Rappe (2019).

adhere to similar processing principles, are deeply integrated, and are in constant communication afforded by lateral connections.

5.3 Negation, predication, and quantification

One important challenge for the proposed account is to provide at least a provisional story about how properties such as negation, predication, and quantification may be realized in the brain with such an architecture. Although a full-fledged answer would merit a separate paper (likely even several), I will attempt to sketch out some promising directions below.

When it comes to predication, there are two important challenges that need to be addressed. The first one concerns extraction of predicates from the PCH as separate representations. The second one relates to the syntactic consistency and expression of predication in language (this distinction is also highlighted in Kiefer 2019). The question of extraction of predicates is not dissimilar to the question of representing relations, and this can be done through the process of conceptualization described in section 4.3. Importantly, conceptual representations do not deal with syntactic consistency, but rather define content, for example, an action or a property (although PCH may inform LH when it comes to coordinating the units of linguistic expression). The details of linguistic expression, such as grammatical coherence, on the other hand, can be resolved purely within the LH. Further, changing specific elements in a proposition may be achieved through the feedback loop between LH and PCH realized by the lateral connections. For a more detailed discussion of predication in the connectionist networks see Kiefer 2019.

The question of quantification also has two aspects. The first aspect is quantification over a finite number of entities, which is represented in language by the vague quantifiers such as “few”, “several”, or “many”. The second aspect is that of the universal quantifiers, such as “all” or “every”, which introduce an additional aspect of infinitude. When it comes to the vague quantifiers, my hunch is that they do not refer to any specific numerical quantities but instead specify contextual and communicative factors such as, for example, the relative size of the objects involved in the scene, or their expected frequency (Moxey and Sanford 1993; Newstead and Coventry 2000; Rakapakse et al. 2005a). Vague quantifiers are then used in language to represent ad hoc, “thin” concepts relaying an estimation of precise information in a context-dependent manner. One possibility is that the use of vague quantifiers is initially learned in the

perceptual domain and then can be applied outside this context. A simple connectionist model of quantification of visual information is presented, for example, in Rakapakse et al. 2005a. The authors report that the networks in the model are able to perform the production of “psychological numbers” produced by human subjects and that, in producing such judgements, they use similar mechanisms (Rajapakse et al. 2005a, b). A provisional story regarding universal quantification is presented in Kiefer 2019 (specifically, section 6.8).

Negation, perhaps, presents the most interesting challenge. A large body of literature suggests that negation makes comprehension more difficult or, at least, takes longer to be processed (see e.g., Fodor and Garrett 1966; Carpenter et al. 1999; Tettamanti et al. 2008; Bahlmann et al. 2011). Sometimes, this is taken as a sign that the processing of utterances with negation requires an additional step compared to those not involving negation. More recently, Yosef Grodzinsky and colleagues (2020) have found that negation is governed by a brain mechanism located outside the language areas, which further suggests that negation may not be a linguistic process. Together with the finding that negative phrases or sentences yield reduced levels of activity in regions involved in the representation of the corresponding affirmative meanings, this supports the decompositional approach to negation—the idea that a positive reversal of the negative utterance is represented first, which is then followed by inhibition of this representation. Further investigation in the nature and levels of brain activity led Liuba Papeo and Manuel de Vega (2020) to conclude that “processing negated meanings involves two functionally independent networks: the response inhibition network and the lexical-semantic network/network representing the words in the scope of negation” (p.741). This naturally aligns with the approach proposed in this paper. Perhaps, negation, at least, at the sentential level, is realized as inhibition of the reverse representation in the PCH. Given lateral connectivity between PCH and LH, as well as simultaneous updating at all levels, it seems plausible that the negation aspect is settled after the main factual content-bearing elements of the utterances are represented.

6. Conclusions

My aim was to show that compositionality and generality may not necessarily present problems for PP even under the assumption of a predictive coding architecture. The proposal outlined

above offers a new perspective on higher-order cognition in linguistic and non-linguistic creatures as well as different types of non-linguistic cognition in humans. Completing such an account would require extensive cooperation of philosophers and cognitive scientists, but my point is theoretical: two abilities are sufficient to capture conceptual thinking in predictive terms—the ability of concept formation as construction of dynamic representational models and a full-fledged linguistic apparatus, also embedded in PP. What is important here is to consider lateral connectivity. Sticking to the very literal sense of top and bottom in the generative hierarchies means looking at generative models as conceptual networks in an overly Fodorian way and ignoring the simultaneous model updating.

The separation and interaction between language and thought raises no shortage of questions, many being already explored. My interest here is how PP provides a twist on these discussions, offering both a new way to ask questions and formulate the answers. Importantly, separating language and thought does not betray the unifying spirit of PP as the systems communicate extensively at different levels, mutually informing and bootstrapping one another. Further, as Williams notes, it is not certain that we need to commit to the single predictive coding architecture within PP at all. Biology and current discussions warn us that a plurality of principles may be at stake. That said, prediction error minimization seems like a useful explanatory concept and if compositionality and generality do not present unresolvable problems for PP in principle, there is no need to get off the PP horse right now, at least when it comes to conceptual thought.

References

- Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Frontiers in Psychiatry*, 4, 47.
- Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and brain sciences*, 33(4), 245–266.
- Bahlmann, J., Mueller, J. L., Makuuchi, M., & Friederici, A. D. (2011). Perisylvian functional connectivity during processing of sentential negation. *Frontiers in Psychology*, 2, 104.
- Barsalou, L. W., & Prinz, J. J. (1997). Mundane creativity in perceptual symbol systems. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (p. 267–307). American Psychological Association.
- Bortolotti, L., & Sullivan-Bissett, E. (2017). How can false or irrational beliefs be useful? *Philosophical Explorations*, 20(sup1), 1–3.
- Call, J., & Tomasello, M. (2011). Does the chimpanzee have a theory of mind? 30 years later. In *Human Nature and Self Design* (pp. 83–96). mentis.
- Carpenter, P. A., Just, M. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1999). Time course of fMRI-activation in language and spatial networks during sentence comprehension. *Neuroimage*, 10(2), 216–224.
- Chater, N., Oaksford, M., Hahn, U., & Heit, E. (2010). Bayesian models of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 811–823.
- Clark, A. (2019). Consciousness as generative entanglement. *The Journal of Philosophy*, 116(12), 645–662.
- Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36(3), 181–204.
- Clark, A. (2000). *Mindware: An introduction to the philosophy of cognitive science*. Oxford University Press.
- Clark, A., Friston, K., & Wilkinson, S. (2019). Bayesing qualia: Consciousness as inference, not raw datum. *Journal of Consciousness Studies*, 26(9–10), 19–33.
- Clark, A., & Wilkinson, S. (2018). *Advanced Topics in Mind, Language, and Embodied Cognition*. [PowerPoint slides].
- Colombo, M. & Hartmann, S. (2017). Bayesian cognitive science, unification, and explanation. *The British Journal for the Philosophy of Science*, 68(2), 451–484.

- Deroy, O. (2019). Predictions do not entail cognitive penetration: “Racial” biases in predictive models of perception. In C. Limbeck-Lilienau and F. Stadler (Eds.), *The Philosophy of Perception* (pp. 235–248). Berlin: De Gruyter.
- Dołęga, K., & Dewhurst, J. E. (2020). Fame in the predictive brain: a deflationary approach to explaining consciousness in the prediction error minimization framework. *Synthese*, 1–26.
- Dove, G. (2018). Language as a disruptive technology: abstract concepts, embodiment and the flexible mind. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752), 20170135.
- Dutilh Novaes, C. (2012). *Formal languages in logic: A philosophical and cognitive analysis*. Cambridge University Press.
- Fedorenko, E., & Varley, R. (2016). Language and thought are not the same thing: evidence from neuroimaging and neurological patients. *Annals of the New York Academy of Sciences*, 1369(1), 132.
- Finn, J. K., Tregenza, T., & Norman, M. D. (2009). Defensive tool use in a coconut-carrying octopus. *Current biology*, 19(23), R1069–R1070.
- Fiorito, G. & Scotto, P. (1992). Observational learning in *Octopus vulgaris*. *Science*, 256(5056), 545–547.
- Fletcher, P. C. & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, 10(1), 48.
- Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. Oxford University Press on Demand.
- Fodor, J., & Garrett, M. (1966). Some reflections on competence and performance. *Psycholinguistic papers*, 135-179.
- Frankish, K. (2018). Inner Speech and Outer Thought. *Inner Speech: New Voices*, 221.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100, 25–50. Reprinted in Frege G. (1956). *The philosophical writings of Gottlieb Frege* (trans: Black M).
- Frege, G. (1879). Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought. *From Frege to Gödel: A source book in mathematical logic, 1931*, 1–82.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2), 127–138.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London: Biological Sciences* 360(1456), 815–836.

- Friston, K. J., & Frith, C. D. (2015). Active inference, communication and hermeneutics. *Cortex*, 68, 129–143.
- Gabriel, G., Hermes, H., Kambartel, F., Thiel, Ch., Veraart, A. (Eds.). (1976). Gottlob Frege. *Wissenschaftliche Briefwechsel*. Felix Meiner: Hamburg.
- García-Carpintero, M. (1996). Two spurious varieties of compositionality. *Minds and Machines*, 6(2), 159–172.
- Gilead, M., Trope, Y., & Liberman, N. (2020). Above and beyond the concrete: The diverse representational substrates of the predictive brain. *Behavioral and Brain Sciences*, 43.
- Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, 65(2-3), 231-262.
- Grodzinsky, Y., Deschamps, I., Pieperhoff, P., Iannilli, F., Agmon, G., Loewenstein, Y., & Amunts, K. (2020). Logical negation mapped onto the brain. *Brain Structure and Function*, 225(1), 19-31.
- Handjaras, G., Ricciardi, E., Leo, A., Lenci, A., Cecchetti, L., Cosottini, M., ... & Pietrini, P. (2016). How concepts are encoded in the human brain: a modality independent, category-based cortical organization of semantic knowledge. *Neuroimage*, 135, 232–242.
- Harman, G. (2015). *Thought*. Princeton University Press.
- Heavey, C. L., & Hurlburt, R. T. (2008). The phenomena of inner experience. *Consciousness and cognition*, 17(3), 798-810.
- Hermes, H., Kambartel, F., Kaulbach, F., Long, P., & White, R. (1981). Gottlob Frege. *Posthumous writings*.
- Hohwy, J. (2020). New directions in predictive processing. *Mind & Language*, 35(2), 209–223.
- Hohwy, J. (2013). *The predictive mind*. Oxford University Press.
- Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, 3, 96.
- Hohwy, J., Roepstorff, A. & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, 108(3), 687–701.
- Horga, G., Schatz, K. C., Abi-Dargham, A. & Peterson, B. S. (2014). Deficits in Predictive Coding Underlie Hallucinations in Schizophrenia. *The Journal of Neuroscience*, 34(24), 8072–8082.
- Hvorecny, L. M., Grudowski, J. L., Blakeslee, C. J., Simmons, T. L., Roy, P. R., Brooks, J. A., ... & Holm, J. B. (2007). Octopuses (*Octopus bimaculoides*) and cuttlefishes (*Sepia pharaonis*, *S. officinalis*) can conditionally discriminate. *Animal cognition*, 10(4), 449–459.

- Janssen, T. M. (2001). Frege, contextuality and compositionality. *Journal of Logic, Language and Information*, 10(1), 115–136.
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
- Kiefer, A. B. (2019). *A defense of pure connectionism*. (Doctoral dissertation, City University of New York, NY). Retrieved from https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=4098&context=gc_etds.
- Kiefer, A., & Hohwy, J. (2018). Content and misrepresentation in hierarchical generative models. *Synthese*, 195(6), 2387-2415.
- Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience*, 8, 302.
- Levinson, S. C. (1997). From outer to inner space: linguistic categories and non-linguistic thinking. *Language and conceptualization*, 13–45.
- Machery, E. (2005). You don't know how you think: Introspection and language of thought. *The British Journal for the Philosophy of Science*, 56(3), 469-485.
- Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, 48(7), 788-804.
- Michel, C. (2020). Concept contextualism through the lens of Predictive Processing. *Philosophical Psychology*, 33(4), 624–647.
- Michel, C. (2016). *What could concepts be in the Predictive Processing framework?* (Master dissertation, the University of Edinburgh, Scotland, UK). Retrieved from <https://era.ed.ac.uk/handle/1842/21875>.
- Miller, M., & Clark, A. (2018). Happily entangled: prediction, emotion, and the embodied mind. *Synthese*, 195(6), 2559–2575.
- Moxey, L. M., & Sanford, A. J. (1993). *Communicating quantities: A psychological perspective*. Lawrence Erlbaum Associates, Inc.
- Nelsen, R. (1993). *Proofs without words: exercises in visual thinking*. Washington, DC: *Mathematical Assoc. of America*.
- Newstead, S. E., & Coventry, K. R. (2000). The role of context and functionality in the interpretation of quantifiers. *European Journal of Cognitive Psychology*, 12(2), 243-259.
- Newton, J. R., & Sur, M. (2005). Rewiring cortex: Functional plasticity of the auditory cortex during development. In *Plasticity and signal representation in the auditory system* (pp. 127–137). Springer, Boston, MA.

Palmer, C. J., Paton, B., Kirkovski, M., Enticott, P. G., & Hohwy, J. (2015). Context sensitivity in action decreases along the autism spectrum: a predictive processing perspective. *Proceedings of the Royal Society of London B: Biological Sciences*, 282(1802), 2014–1557.

Papeo, L., & de Vega, M. (2020). The neurobiology of lexical and sentential negation. In V. Déprez & M.T. Espinal (Eds.), *The Oxford Handbook of Negation*. Oxford University Press, USA.

Penny, W. (2012). Bayesian models of brain and behaviour. *ISRN Biomathematics*, 2012.

Pellicano, E. & Burr, D. (2012). When the world becomes ‘too real’: a Bayesian explanation of autistic perception. *Trends in Cognitive Sciences* 16(10). 504–510.

Rajapakse, R. K., Cangelosi, A., Coventry, K. R., Newstead, S., & Bacon, A. (2005a, September). Connectionist modeling of linguistic quantifiers. In *International Conference on Artificial Neural Networks* (pp. 679-684). Springer, Berlin, Heidelberg.

Rajapakse, R., Cangelosi, A., Coventry, K., Newstead, S., & Bacon, A. (2005b, April). Grounding linguistic quantifiers in perception: Experiments on numerosity judgments. In *2nd Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*.

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.

Rappe, S. (2019). Now, Never, or Coming Soon? Prediction and Efficient Language Processing. *Pragmatics & Cognition*, 26(2-3), 357-385.

Richter, J. N., Hochner, B., & Kuba, M. J. (2016). Pull or push? Octopuses solve a puzzle problem. *PloS one*, 11(3).

Russell, S., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed.). London: Pearson.

Seth, A. K. (2015). The Cybernetic Bayesian Brain—From Interoceptive Inference to Sensorimotor Contingencies, (w:) Open MIND, red. T. Metzinger, JM Windt.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11), 565–573.

Sorensen, R. A. (2004). *Vagueness and contradiction*. Clarendon Press.

Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A., Paliwal, S., Gard, T., ... & Petzschner, F. H. (2016). Allostatic self-efficacy: a metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in human neuroscience*, 10, 550.

Tacca, M. C. (2011). Commonalities between perception and cognition. *Frontiers in psychology*, 2, 358.

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.

Tettamanti, M., Manenti, R., Della Rosa, P. A., Falini, A., Perani, D., Cappa, S. F., & Moro, A. (2008). Negation in the brain: Modulating action representations. *Neuroimage*, 43(2), 358-367.

Thagard, P. (2019). *Brain-Mind: From Neurons to Consciousness and Creativity (Treatise on Mind and Society)*. Oxford University Press.

Tversky, B. (2019). *Mind in motion: How action shapes thought*. Hachette UK.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eysen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, 121(4), 649.

Vance, J. (2015). Review of The Predictive Mind. *Notre Dame Philosophical Reviews*.

Velasco, P. F., & Loev, S. (2020). Affective experience in the predictive mind: a review and new integrative account. *Synthese*, 1-36.

Wiese, W., & Metzinger, T. (2017). Vanilla PP for philosophers: A primer on predictive processing. In T. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing*: 1. MIND Group, Frankfurt am Main.

Wilkinson, S., & Fernyhough, C. (2018). *When Inner Speech Misleads*. Oxford University Press.

Williams, D. 2018. Predictive coding and thought. *Synthese*. 1–27.

Zuberbühler, K. (2020). Syntax and compositionality in animal communication. *Philosophical Transactions of the Royal Society B*, 375(1789), 20190062.