Behavioral and Brain Sciences

cambridge.org/bbs

Above and beyond the concrete: The diverse representational substrates of the predictive brain

Michael Gilead[a], Yaacov Trope[b] and Nira Liberman[c]

[a]Department of Psychology and Zlotowski Center for Neuroscience, Ben-Gurion University, Beersheba 84105, Israel; [b]Department of Psychology, New York University, New York, NY 10003 and [c]Department of Psychology, Tel-Aviv University, Tel-Aviv 69978, Israel.
mgilead@bgu.ac.il
yaacov.trope@nyu.edu
niralib@tauex.ac.il
http://www.gileadlab.net/
http://www.psych.nyu.edu/tropelab/
https://en-social-sciences.tau.ac.il/profile/niralib

Target Article

Cite this article: 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**, e121: 1–74. doi:10.1017/S0140525X19002000

Target Article Accepted: 30 June 2019
Target Article Manuscript Online: 18 July 2019
Commentaries Accepted: 23 October 2019

Keywords:
abstraction; concepts; construal level theory; declarative memory; mental representation; mental simulation; prediction; predictive processing; reinforcement learning; symbolic thought; theory of mind

What is Open Peer Commentary? What follows on these pages is known as a Treatment, in which a significant and controversial Target Article is published along with Commentaries (p. 19) and an Authors' Response (p. 53). See bbsonline.org for more information.

Abstract

In recent years, scientists have increasingly taken to investigate the predictive nature of cognition. We argue that prediction relies on abstraction, and thus theories of predictive cognition need an explicit theory of abstract representation. We propose such a theory of the abstract representational capacities that allow humans to transcend the "here-and-now." Consistent with the predictive cognition literature, we suggest that the representational substrates of the mind are built as a *hierarchy*, ranging from the concrete to the abstract; however, we argue that there are qualitative differences between elements along this hierarchy, generating meaningful, often unacknowledged, *diversity*. Echoing views from philosophy, we suggest that the representational hierarchy can be parsed into: *modality-specific* representations, instantiated on perceptual similarity; *multimodal* representations, instantiated primarily on the discovery of spatiotemporal contiguity; and *categorical* representations, instantiated primarily on social interaction. These elements serve as the building blocks of *complex structures* discussed in cognitive psychology (e.g., episodes, scripts) and are the inputs for mental representations that behave like functions, typically discussed in linguistics (i.e., *predicators*). We support our argument for representational diversity by explaining how the elements in our ontology are all required to account for humans' predictive cognition (e.g., in subserving logic-based prediction; in optimizing the trade-off between accurate and detailed predictions) and by examining how the neuroscientific evidence coheres with our account. In doing so, we provide a testable model of the neural bases of conceptual cognition and highlight several important implications to research on self-projection, reinforcement learning, and predictive-processing models of psychopathology.

Recent years have seen the emergence of a wave of influential theories that highlight the predictive nature of cognition – the so-called *predictive brain* framework (see Bar 2011; Clark 2013). A common denominator of these theories is that they paint a picture of the mind wherein our mental representations of the world become active *before* we engage with reality (i.e., so-called "top-down" processing); this view contrasts with traditional perspectives that assumed that our representation of the current state of the world emerges only *after* we have acquired evidence from our sense organs (i.e., "bottom-up" processing).

A prominent theory within this framework is the *Predictive Processing* (PP) approach (e.g., Bar et al. 2006; Friston 2005). Proponents of PP argue that every encounter we have with reality is akin to scientific hypothesis testing. For example, a person who is about to open the fridge already has prior representations of what they are about to see (e.g., a milk carton); to the extent that this representation successfully predicted the event to come, there is no need for much additional cognitive processing; however, when a discrepancy between the prior representation and bottom-up inputs is detected (i.e., if there is no milk left, a so-called *prediction error*), then there is a need to update the mental representation in light of the new evidence. Often, theories within the PP camp argue that such updates mimic rules of normative reasoning and, specifically, Bayesian inference.

The ideas of the PP approach have long been influential in the domain of perception science (e.g., Gregory 1980; Helmholtz 1860/1961) and have become paradigmatic in the neuroscientific literature on this topic (e.g., Bar et al. 2006; Friston 2005; Rao & Ballard 1999). More recently, generalizations of PP theory to the domain of action (most notably, active inference theory further described in section 3, e.g., Friston et al. 2009) have been able to account for diverse phenomena in areas such as decision-making (e.g., Schwartenbeck et al. 2013) and psychiatry (e.g., Barrett et al. 2016; Friston et al. 2014; Powers et al. 2017). Some have argued that the PP approach, and specifically Active Inference Theory, may provide a unified theory of brain function (Hohwy 2013) and a grand paradigm for cognitive science (Clark 2013).

Another recent group of influential theories in the "predictive brain" camp may not subscribe to an all-encompassing role for predictive mechanisms – but nonetheless, ascribe a critical role to prediction. Specifically, these approaches stress that cognition greatly relies on *prospection* (or future-oriented *mental time travel*; Suddendorf & Corballis 2007; Tulving 1984) – the ability to deliberately create explicit representations of future events, and use these representations to guide behavior. Like PP theory, research on prospection has been highly influential across numerous domains of investigation. Within memory research, research on prospection argues that the function of declarative memory is in enabling simulation of future events (e.g., Schacter et al. 2007). Within the Reinforcement Learning (RL) literature, research on prospection investigates how organisms often make decisions by relying on so-called *model-based* algorithms (e.g., Daw et al. 2011) that explicitly simulate future outcomes (e.g., Redish 2016). Within comparative psychology, research on prospection suggests that the evolutionary success of humans[1] stems from our capacity for future-thought (e.g., Suddendorf 2013).

It is commonly held that humans' advanced ability for prospection must rely on similarly advanced representational capacities. Broadly speaking, these representational abilities are referred to as the ability for *abstraction* or as *abstract thought* – and are contrasted with more *concrete thought*. This distinction between relatively abstract and concrete mental representation has been integral to theories of the predictive brain. For example, the literature on mental time travel (e.g., Schacter et al. 2007) argues that we are able to imagine and predict future events by relying on the abstract representational capacities afforded by declarative memory (specifically, episodic memory[2]) as opposed to procedural/non-declarative memory. In the RL literature, the concrete-abstract dimension is reflected in the distinction between *model-free* learning that relies on relatively simple associations (e.g., Schultz et al. 1997), and *model-based* processing that relies on hierarchical, structured *cognitive models* (Tolman 1948) of potentially complex *state-spaces* (e.g., Daw & Dayan 2014). Within PP theory, *higher-level*, more abstract units in a representational hierarchy form predictions that inform and interact with *lower-level* units.

Michael Gilead, is an Assistant Professor of Psychology at Ben-Gurion University of the Negev. His research utilizes neuroimaging, experimental, and big-data methodologies to investigate humans' symbolic cognition and symbolic interaction. He has authored papers on topics such as mentalizing, mental simulation, language and cognition, emotion, and decision making. He is the recipient of the APS Rising Star award.

Yaacov Trope is a Professor of Psychology at New York University. He is the author of over 150 publications in the general areas of social cognition, motivation, and self-regulation. He is a member of the American Association for Arts and Sciences, Fellow of the Association for Psychological Science, and a recipient of the Society for Personality and Social Psychology Career Achievement Award, the Association for Psychological Science Mentor Award for Lifetime Achievement, and the Tom Ostrom Award in Social Cognition.

Nira Liberman is a Professor of psychology at Tel-Aviv University. She is the author of over 120 publications on motivation, social cognition, decision-making, and Psychopathology. She is Fellow of the Association for Psychological Science and the recipient of a career trajectory award of the Society for Experimental Social Psychology career trajectory award and the Tel Aviv University Kadar family award for excellence in research.

In light of the centrality of abstract mental representations in theories of the predictive brain, an in-depth account of representational abstractness seems essential for developing a comprehensive account of predictive cognition. The present article aims to provide such an account. Luckily, this work does not have to start from scratch. Decades of research on higher-order cognition have generated rich and intricate theoretical conceptualizations of the many different representational entities that give rise to abstract thought. We believe that, to date, this richness and intricacy has not been sufficiently tied to the newly evolving paradigm of the predictive brain and its neural substrates. Possibly, the lack of integration between cognitive science's rich past and its present hinders future development.

This dissociation can be traced to an earlier problem of insufficient integration across different theories of abstract cognition, and a lack of a joint vocabulary across different influential frameworks. In the current manuscript, we provide a unified conceptualization of abstraction, which we further integrate into the newly evolving framework of the predictive mind. *In doing so, we aim to generate a comprehensive account of the representational bases of people's ability to traverse the here and now.* Our account, which evolved from previous research on Construal Level Theory (e.g., Liberman & Trope 2008; 2014; Trope & Liberman 2010) aims to achieve two goals. First, we attempt to shed light on the diversity of abstract mental representation; second, we wish to integrate this diversity under a unified framework that could be tested, refined and revised through further experimentation.

We begin by suggesting a definition of the act of abstraction and describing how it relates to the different representational entities that have been designated as "abstract" in previous literature (sections 1–2). We then provide our account of the process of "mental travel" (sect. 3) and explicate how the different abstract representational entities expand humans' mental travel ability (sect. 4). Finally, we discuss how the current framework can provide a theoretical basis for understanding the neural substrates of the predictive brain (sect. 5).

1. How do abstract mental entities emerge?

Despite the wide use of the term "abstraction," there have been few attempts to provide a definition of this fundamental construct. We begin by providing a definition that will serve as the basis for our subsequent analysis of abstract mental representations and their role in prediction.

1. Our definition focuses on abstraction as a phenomenon that pertains to mental states[3] (i.e., beliefs, desires, intentions). These mental states are "directed at" a (physical or mental) object, and have conditions of satisfaction[4,5] – for example, when a person believes[6] that the earth is round, then this belief can be satisfied when novel observations suggest that this is

https://doi.org/10.1017/S0140525X19002000

indeed the case; when a person desires ice cream, then this desire can be satisfied if she eats ice cream; when a person intends to file her taxes, then this intention can be satisfied once the taxes are filed. In other words, satisfaction is akin to minimizing a discrepancy between one's internal state and the state of the world.[7]
2. Whenever there exists a mental state that is satisfiable by object A, but not by object B, these objects can be said to exist as subjectively distinguishable objects in the mind of the perceiver.
3. We define the act of abstraction as the formation of a belief[8] that two or more subjectively distinguishable objects satisfy a belief, a desire or an intention.[9]

For example, forming a belief that both jogging and dieting would satisfy my intention to lose weight is an act of abstraction. Here are a few additional examples:

1. An infant desires milk. An instinct causes it to put various objects in its mouth. A perceptual pattern, which – from an external perspective – we call "mother," repeatedly satisfies the infant's desire, regardless of whether mother is wearing a (tickling) sweater or a (smooth) T-shirt. According to our definition, *once this substitutability is represented as a new entity in the infant's mental system* (once it forms the belief that the distinguishable objects satisfy the desire) it has performed an act of abstraction. Such acts of abstraction are often discussed under the term *generalization* (e.g., Pearce 1987; Shepard 1987) and *recognition,* in research into basic learning processes.
2. A rat learns that pressing a lever in a red cage is substitutable with pressing a lever in other red cages (where it produces a reward), but not with pressing a lever in a blue cage. Such scenarios are discussed under the term *situation/state recognition* (e.g., Gershman et al. 2010; Redish et al. 2007) in the RL literature.
3. A child is told by her parents "look at the kitty" whenever they encounter a cat, and forms the belief that various hairy, four-legged creatures are substitutable in satisfying the belief "(this) is a kitty." Such processes are treated under the heading *concept/category/word formation/learning/acquisition* (e.g., Bloom & Markson 1998; Carey 2009; Kruschke 1992; Tenenbaum et al. 2011).
4. An interviewer forms the implicit belief that a job applicant is predominantly similar to other candidates from the same ethnic background, rather than to other candidates with similar work experience. Such acts are discussed under the heading *social stereotyping* in the social-psychological literature (e.g., Banaji et al. 1993; Zarate & Smith 1990), and can be seen as a specific case of *categorization* or *similarity judgment/analogical comparison* (e.g., Gentner 1983; Medin et al. 1993).

All these examples highlight the commonality of different acts of abstraction. Just like the act of prediction, abstraction appears as an omnipresent regularity of the mind – and a defining property of cognition. However, it must be stressed that this description of abstraction pertains to the computational level of analysis (Marr 1982), namely, it is an (abstract) description that does not correspond to one specific mechanism or neural hardware.[10]

It is also important to note that forming the belief that two distinguishable objects satisfy a mental state does not mean that in order to perform an act of abstraction one must consider two particular exemplars. Although it is possible to perform an act of abstraction in a bottom-up manner, by having a thought like: "both object A and object B give rise to a sensation of tastiness," it is clearly also possible to perform the act of abstraction by relying on the outputs of previous acts of abstraction, for example, by having a thought such as: "things that are made of chocolate give rise to a sensation of tastiness" and "things that are tasty are often made of chocolate." Importantly, in all cases, abstraction deems (at least) two subjectively distinct objects as equivalent – as well as any additional objects that would satisfy the same mental state.

1.1. The outputs of the act of abstraction

As follows from the definition, the output of abstraction is a belief that two or more subjectively distinguishable objects satisfy a belief, a desire or an intention. The emergent output of this belief is that a person possesses a mental representation that allows them to associate between the objects and the mental state that they satisfy.

1. We refer to the set of distinguishable objects as the concreta.[11]
2. We refer to the rule (or algorithm, function) that determines/picks-out the set of equivalent objects (the concreta) for a given mental state as the criterion of substitutability[12] (a notion related to a *sense* in the Fregian theory, i.e., Frege 1892/1952b).

As noted above, criteria of substitutability can take a form such as "things that are tasty are often made of chocolate"; this means that they can implement a *theory* (Murphy & Medin 1985). Theories allow us to *generate predictions of (or imagine) future members of a set*, rather than just assign a probability of class membership given a list of features (i.e., they implement a *generative model*, Ng & Jordan 2002).

3. We refer to the newly generated mental object that instantiates (1) and stands for (2) as the abstractum.
4. Because cognition allows the outputs of abstraction to serve as concreta for additional acts of abstraction (Berwick et al. 2013), we can speak of mental representations as forming a continuum of abstractness. We define abstractness as a relative term that refers to the relation between two abstracta. Whenever we can say that abstractum X is part of the concreta of abstractum Y, we will say that abstractum Y is more abstract than abstractum X.

In our definition, if two objects are not distinguishable in any sense by the observer, then there is only a single object in mind, hence there is no abstraction. The requirement of distinguishability means that abstraction involves having at least two dimensions[13] in mind: one dimension on which the stimuli differ, and another dimension on which they will be considered identical. Thus, when performing an act of abstraction, one makes a (conscious or non-conscious) decision[14] on which dimension is central, and by doing so, one designates other dimensions as secondary or irrelevant in the current context (see Shapira, Liberman et al. 2012, for a related definition of abstraction).

Because abstraction entails selecting/attending to one dimension and disregarding other dimensions that might be salient, many acts of abstraction likely rely on cognitive operations often referred to as "cognitive control" and "selective attention" (e.g., Botvinick et al. 2001; Macleod 1991). The degree to which cognitive control is needed for an act of abstraction depends on

https://doi.org/10.1017/S0140525X19002000

the relative salience of the (attended) dimension of substitutability versus the (ignored) dimension on which the concreta subjectively differ (see Deacon 1997, for a related idea).

Consider a psychology experiment wherein a child sees a picture of a person who has a mustache and wears a dress. The child can form the belief that the person is substitutable with other males (because of the mustache) or with other females (because of the dress). What made "clothing type" a relevant and salient dimension for the child? (see Fodor 1998; Goodman 1972).[15] Clearly, the dimensions used for generating new criteria of substitutability come from our prior network of beliefs (Murphy & Medin 1985). However, explaining the formation of new beliefs merely based on prior beliefs leads to an infinite regress. The dimensions that form our criteria of substitutability must have been originally introduced into our mind at some point.

The question of where these dimensions come from is a major topic of investigation in the field of developmental psychology and concept formation. Broadly speaking, these dimensions can be acquired in three ways: they can be innate (e.g., infants may have an innate biological mechanisms that determine that the velocity and direction of an object is an important dimension to attend to in the newly-discovered world), shaped by personal experience and the gradual discovery of statistical regularities (e.g., a rat may discover that a specific auditory cue is a useful dimension along which to group outcomes), or transferred from other people (e.g., a child may learn from society that gender is a meaningful dimension along which to categorize people).

2. The diverse representational ontology that cognitive science should not ignore

As noted earlier, the ability to use abstracta as inputs for further acts of abstraction generates a continuum of abstractness. However, we argue (and in sect. 5, review evidence) that there are qualitative distinctions along this continuum. Performing the act of abstraction by relying on different types of inputs and different types of dimensions gives rise to qualitatively different types of abstracta.

2.1.1. Modality-specific features, objects, and relations.[16]

The first steps in moving beyond a concrete representation can be traced to the discovery of identity between different perceptual features (e.g., color, loudness, a nose), and the formation of representational permanence of objects on the basis of *perceptual pattern similarity* (or "object permanence"; Piaget 1954). Object permanence relies on the formation of a belief that different sensory impressions that appear across different spatial and temporal contexts are equivalent, in the sense that they all pertain to the same object (e.g., "the image of my dog"). Likewise, perceptual patterns which are best described as relations (e.g., sound A is louder than sound B; objects A and B are similar in color; e.g., Martinho & Kacelnik 2016) can perhaps be represented in a modality-specific manner, and be deemed equivalent across different instantiations. *Such modality-specific abstracta probably often rely on different innate, hard-wired dimensions such as pitch and color* (Baillargeon et al. 1985; Carey 2009).

2.1.2. Multimodal features, objects, and relations[17]

Once individuals generate some set of modality-specific abstracta, they may use them in further acts of abstraction. This is because percepts that *cannot be grouped together based on perceptual pattern similarity* may also be deemed as substitutable – whenever they are *bound together by spatiotemporal contiguity*. For example, when a toddler experiences the (modality-specific) sound of a dog and the (modality-specific) sight of the dog at the same time, it generates the multimodal abstractum of the dog. As another example, an animal that repeatedly hears a bell before it gets food can bind these two patterns together. *Whereas the discovery of modality-specific abstracta often relies on innate substitutability criteria, multimodal abstracta are often acquired via personal experience.*

Two types of multimodal representations warrant special consideration. First, a person might group together various different objects that share a temporal context (e.g., "the dog chasing the Frisbee in the park.") The resulting abstractum can be called a mental episode[18] (Tulving 1984). A second important class of multimodal representations are lemmas, which are the entities of our mental lexicon (Roelofs 1992). These are the abstracta whose concreta include modality-specific representations of a feature/object/relation as well as the linguistic sign (written, spoken) that co-occurs with it. For example, the lemma of a banana will group images of bananas, taste experiences associated with bananas, alongside with the visual symbol "banana," and the auditory pattern "ba-na-na."

2.1.3. Categories

Multimodal abstracta can be used as building blocks for further acts of abstraction – even in situations where they *cannot be grouped together based on spatiotemporal co-occurrence* (Murphy & Medin 1985). This results in a *category*.[19] Categories arise, for example, when we form the belief that different multimodal abstracta such as bats and whales are all substitutable as mammals. Even the brightest child will probably be unable to acquire the category "mammal" (that includes both bats and whales) without being taught by someone. *In other words, many categories probably rely on socially acquired dimensions and substitutability criteria that are subserved by lemmas.*

Categories are often discussed within the context of perceptions and objects but clearly extend to actions and relations (see Bargh 2006; Hommel et al. 2001; Weingarten et al. 2016, for evidence of category-based organization of actions). For example, when we form the belief that different relations between people (e.g., "giving change to a beggar" and "giving a gift to a friend") are substitutable acts of "giving" we have generated a categorical representation of action (Semin & Fiedler 1991; Vallacher & Wegner 1987; see Mahler 1933 for initial research on substitutability in action categories).

An especially important type of category is one whose concreta are intangible entities. As noted, basic acts of abstraction involve transforming a percept of a particular spatiotemporal event into a modality-specific object (e.g., *the image of my* bicycle); forming a multimodal representation is another step (*my* bicycle), which can be followed by the formation of a category which includes different multimodal objects (the category "methods of transportation"). However, even non-particular (i.e., "universal") objects ("methods of transportation") and relations (e.g., "behind") can refer to entities whose particular instantiations are detected by the senses. Yet, categorical abstracta can pertain to objects that are undetected by our senses, for example, "energy," and "future." These objects are often referred to as "abstract objects"; this terminology, however, is confusing, because, as we explained above, seemingly "concrete objects" such as "bicycle" are also the product of abstraction.

Whenever two or more intangible entities are believed to satisfy a mental state, we will refer to the resulting abstractum as an

https://doi.org/10.1017/S0140525X19002000

intangible abstractum. The big question surrounding intangible abstracta is how intangible objects, features, and relations take their place in the mind in the first place; this ability has been suggested as the crucial point of divergence between humans and non-human animals (Penn et al. 2008). The question of how intangible features are initially "perceived" is beyond the scope of our current discussion. However, broadly speaking, it seems safe to assume that although some intangible dimensions may have an *innate* basis, or may be emergent properties discovered via *personal experience* (using socially-unmediated statistical learning mechanisms), many intangible dimensions (e.g., using zodiac signs as a criterion according to which to predict romantic success) are likely transmitted from mind to mind via the use of lemmas and words and as such rely on the pre-existence of multimodal abstracta (i.e., they are *socially acquired*).

Moreover, it should be noted that the discovery and use of many intangible abstracta are likely especially difficult, in that it requires us to disregard dimensions of equivalence that are typically salient and important. For example, the idea that hateful rhetoric, as well as peaceful rhetoric, may be deemed as identical – in that they are protected under the principle of "free speech" – requires that we ignore dimensions that are typically deep-seated in our mind (the malevolence/benevolence of people).

2.2. Forming complex mental structures

We have broadly outlined how abstraction generates modality-specific abstracta, how these modality-specific abstracta can serve as the concreta of multimodal abstracta – that, in turn typically serve as the concreta for categories. By performing recursive acts of abstraction, humans organize these representations within larger-scale structures (e.g., Smith 1998), mostly discussed in the literature on semantic and episodic memory (e.g., Tulving 1986).

2.2.1. Network structure

The assumption that when we create an abstractum we retain the association between its constituent elements entails that our representations will gradually form a network/graph structure (e.g., Hintzman 1986). Regularities in the world entail that the same abstracta will be formed repeatedly, which in turn implies that some abstracta will have a greater degree of association with each other than others. For example, if the abstractum "cake" is repeatedly categorized as "dessert" (rather than, "baked goods") – and if the intention to eat dessert is repeatedly satisfied by cakes (rather than "chocolate") – cake will become a prototypical concretum of dessert. Namely, it should be processed faster, will be more likely and more confidently retrieved when prompted with "dessert," and will be judged as a better instantiation of this category than chocolate (Rosch & Mervis 1975).

2.2.2. Hierarchical structure

Once we decide (or are told) that for some purpose, *dogs* and *chimpanzees* are substitutable, we can create (for example) the abstractum *mammal.* We can then continue and decide that *mammals and fish* are also substitutable (being *vertebrates*), and so forth. In doing so, we gradually create a hierarchical system of mental representations that are organized in a tree-like structure of many-to-one relations (e.g., Collins & Quillian 1969).

Likewise, in the domain of actions, when deciding, for example, that *swimming* and *running* are substitutable as a means to achieve the goal of exercising, and that *exercising* and *eating healthy* are substitutable as means to achieve the goal of maintaining health, one gradually builds *action hierarchies* which serve as the basis of directing and representing goal-directed behavior (Carver & Schier 2001; Cohen 2000; Vallacher & Wegner 1987).

2.2.3. Temporal structure

Earlier we defined *episodes* as a type of multimodal abstracta that bind abstracta based on close temporal contiguity. This may give the impression that episodic memory is a fragmented tapestry of discrete events. However, discrete episodes may also be bound together by spatiotemporal contiguity (e.g., Tulving 1985), thereby generating a hierarchical structure of multimodal abstracta that can extend across hours or even days. For example, episode A (going to the gym) and episode B (taking a shower) may be bound together into episode C (gym and shower), which may be bound together with later episode D, and so forth (see Corballis 2014 for a discussion of the recursive structure of episodic memory).

Episodes that occur at temporally distant contexts (i.e., do not share temporal contiguity) may also be grouped to generate a type of category often called a script (Schank & Abelson 1975). For example, various episodes of airport visits can be deemed as equivalent, generating an airport script. Scripts can likewise be organized within increasingly abstract action categories ("going to the airport," "traveling to a different country," "traveling"), as described earlier.

2.3. How can mental structures interact with other mental structures?

The final entity we posit in our ontology (and which has been widely discussed in linguistics, but rarely in psychology and neuroscience) is the predicator. Like the other types of abstracta we have discussed, predicators instantiate a rule that determines the set of entities that are equivalent in satisfying a particular mental state (the criterion of substitutability). For example, the predicator "red" defines a certain visual-processing property as the dimension along which stimuli are deemed as equivalent, ignoring dissimilarity on other dimensions (such as object identity; its concreta contain different objects like "red dog" and "red Corvette"). Crucially, in order for the abstractum "red" to be a predicator (rather than being the "ordinary" category "red") it must call for the specification of a subset of its concreta – by taking another entity (a different abstractum; e.g., "dog," "apple") as an input argument. As such, predicators are *representations that behave like functions*.

The theoretical distinction between "regular" abstracta and predicators can be traced to Frege (1892/1952a) who famously noted the difference between "saturated" abstract entities (or, in his terminology, "objects") – that do not need other entities in order to be functional, and "unsaturated" abstract entities (or "concepts") – that rely on input objects in order to function. The existence of predicators in our minds is believed to be reflected in (overt) language use[20] (e.g., Pinker 2007). For example, in order for verbs to be functional they require the specification of a noun phrase as an argument (e.g., "lost" is not a meaningful utterance – until you specify who lost what, such as "Cleveland *lost* the championship").

The interdependence of a predicator and its argument means that the operation of predicators (or *predication*) is best described as an interaction; moreover, according to the view presented herein, this interaction is characterized by an asymmetrical

https://doi.org/10.1017/S0140525X19002000

relationship of the predicator and its argument[21] – the "saturated" abstractum, which (prior to its encounter with the predicator) was perfectly fine with denoting a wide variety of concreta (e.g., "dog"), is forced by the predicator into denoting a more limited set (i.e., it no longer denotes any dog, but specifically, a "red dog"). Thus, predicators can be thought of as "concretization machines," in that they modify their input argument in a specific manner.

This specification that predicators entail plays a crucial role in human cognition: it allows us to modify representations in a systematic, rule-based/algorithmic manner (Bogdan 2009; Fodor & Pylyshyn 2014). In doing so, predicators enable a purported "language of thought" (Fodor 1975) – a platform upon which we use mental representations to systematically orchestrate the modification of mental representations.

The computational process whereby predicators modify their arguments has long posed a challenge to theories of cognition (Fodor & Pylyshyn 2014). Most notably, theories of concepts that assume that the act of abstraction entails grouping together sets of particular exemplars (i.e., "exemplar theories"; e.g., Medin & Schaffer 1978) or generating some statistical summary representation of these exemplars (i.e., "prototype theories"; Rosch & Mervis 1975) may struggle[22] to explain how predication can generate previously never-encountered and non-typical objects (e.g., "a smoking caterpillar"; see Fodor & Pylyshyn 2014 for a discussion). In contrast to exemplar and prototype theories, we argued that abstraction requires applying a *criterion of substitutability* that can serve as a generative model for finding instantiations that will satisfy a particular mental state, whether previously encountered or not. According to such a view, predication might be best seen as a type of mental algebra that entails applying the criterion of substitutability of the predicator on top of the criteria of substitutability of its argument (e.g., "tasty cockroaches" are "insects that run on the ground" and also "make you feel good when you eat them").[23]

Some predicators have more than one argument (Marcus 2001). For example, the predicator *lifted* requires that you provide it with an argument for the agent (e.g., *Dana*) and for the patient (e.g., *the dog*). If given proper inputs, *Lifted*(*Dana*, *the dog*) will instantiate a specific relation between them (i.e., the dog is in the air, Dana is not). In this way, a predicator that operates upon more than one argument (a relational predicator, see Doumas et al. 2008; Gentner & Markman 1997; Halford et al. 2010; Marcus et al. 1999; Markman & Stilwell 2001) modifies its arguments by specifying their relation to each other.

Of special importance are intangible relational predicators. They are distinct from representations of predicators such as *Lifted*, because, as their name suggests, they denote a relation that does not have specific perceivable instantiations (for example, "A thinks B" and "A is like B"). Relational predicators likely play an especially important role in higher-order cognition. For example, mastering the use of a system of logical relations allows for the emergence of the formal systems of reasoning (Evans 2003; Goodwin & Johnson-Laird 2013; Kuczaj & Daly 1979). This ability is possibly subserved by a toolkit of intangible relational predicators such as *if*(A,B), *cause*(A,B), *or*(A,B), that designate the specific intangible deontology between the arguments.

Just like in the case of logical relations, the intangible relational predictor that determines that A *symbolizes* B may be of special importance. This *Symbolize*(A,B) predicator allows us to designate the specific intangible relation between a symbol (e.g., a written word) and its referent, B (i.e., "A can replace B in the mental world or in communication, but not in the real world"). This predicator might play a crucial role in humans' ability to safely manipulate mental objects (now symbols), without worrying about manipulating real objects (DeLoache 2004; Leslie 1987), a point to which we return later (see sect. 4). Relatedly, the ability to designate that a representation is explicitly about other mental representations (a "meta-representation," Pylyshyn 1978) may have crucial importance (Leslie 1987). For example, the predicator *Believe*(A,B) is impervious to who is believing, and what is believed – but designates the complex relation between the thing that is believed, the mind of the believer, and the world. As we discuss later, such predicators may be crucial for social cognition.

2.5. Interim summary: Sections 1–2

In sections 1–2 we proposed an ontology of abstract mental representations, based on extant theorizing and research. If cognitive scientists were right in assuming the existence of (at least some) of these entities, then any theory of the mind – theories of the predictive brain included – may need to integrate this representational diversity into their conceptualization. As we will argue in section 4, this diversity seems crucial in order to account for humans' capacity to "traverse the here and now." However, before going into this explication, we must first unpack what this capacity entails.

3. Interlude – What is Mental Travel?

Millions of years of evolution have ingrained within us mechanisms for acquiring and using information. For example, without resourcefulness and effort on our behalf, we are born with nerve endings that detect extreme heat, and the motor reflexes that tell us to distance ourselves from the source of heat. However, although our innate senses supply us with access to quite a lot of information, there is clearly much more knowledge in the world than what meets the eye (or other sense organs).

A proto-human would have had better chances to reproduce, if one day she would have woken up with the knowledge that: 10 kilometers up-north there is a waterfall; tomorrow, boars will visit it; she has a 40% chance of catching one; her friends will want to steal her prey. Such knowledge, however, oftentimes remains obscure, as it cannot be drawn from one's direct experience. Within Construal Level Theory (e.g., Liberman & Trope 2008; 2014; Trope & Liberman 2010), the aforementioned epistemic barriers are termed "dimensions of psychological distance." Trope and Liberman posit that much of the ignorance we face in life is a result of spatial, temporal, and social distance, and because of uncertainty concerning the ontological status of things (hypotheticality) – and that the attempt to mitigate ignorance and uncertainty is an important force in human cognition.

The crucial importance of mitigating uncertainty is also a hallmark of PP theory. Most notably, Active Inference Theory (Friston 2010) subsumes all cognitive activity under a single epistemic imperative – the attempt to reduce the *surprise*[24] that we *expect* to experience in our next interaction with reality (a process termed "free-energy"[25] minimization). In information-theoretic terms, *expected surprise* is the same as uncertainty; thus, Active Inference Theory suggests that every action an organism makes is an attempt to reduce uncertainty.

Several key insights emerge from this account. A first insight (which echoes discussions of the *intentionality* of the mind;

https://doi.org/10.1017/S0140525X19002000

Brentano 1874; for example, Searle 1983; Velleman 1992) is that prior representations/expectations can become consistent with reality via two substitutable routes: organisms can either update these representations so that they cohere with sense data, or act in a manner that alters the world (and makes these predictions accurate). A second key insight is that the epistemic imperative can explain supposedly non-epistemic (i.e., utilitarian) phenomena such as mate-seeking and defensive behaviors. This is because organisms have prior representations/expectations that they will continue to live (e.g., will eat, will not be predated)[26]; given these predictions, and given the substitutability of belief and action, the imperative "to be right" brings about a state wherein predictions are often fulfilled, and organisms survive.

Other theories in the predictive brain landscape (e.g., Suddendorf & Corballis 2007) may not endorse the idea according to which the mitigation of ignorance and uncertainty is *the* goal of the mind, but nonetheless, stress that the ability to predict the future is what led to humans' evolutionary success. Simply stated, these theories argue that the person who was just about to go into a bear's cave, but suddenly gained a glimpse into the unfortunate outcome of such an action, would have outlived his future-myopic friend. Although our account is consistent with theories of prospection, we wish to highlight that futurity is just one of a number of epistemic barriers that humans may face. As described earlier, other dimensions of psychological distance (i.e., spatial distance, social distance, hypotheticality) also engender uncertainty.

Moreover, many of the unknowns of reality do not stem directly from divergence from one's own experience of the here and now. For example, one can observe lightning hit the ground without knowing what caused it, even though this occurrence is not distant in time, space, and social perspective. In other words, understanding the cause of a phenomenon (e.g., by developing a theory of electricity), which affords additional advantages (in the form of artifacts such as electrical lighting), is not reducible to traversing one of the psychological distance dimensions.

Finally, it should be stressed that a central aspect of the uncertainty faced by humans also extends to social reality. Traversing social distance may concern the mental states of individuals (i.e., perspective-taking/mentalizing/theory-of-mind; e.g., Baron-Cohen et al. 1985; Heider 1958; Jones & Davis 1965; Kelley 1973; Malle & Holbrook 2012; Trope 1986;) or may concern social constructions (for example, in attempting to understand the distinction between homicide in the first vs. second degree). Moreover, even when the object of a person's inquiry does not appear to be social in nature (e.g., what is the shape of the earth), people might perceive other minds to be of much relevance, and wish to align themselves with the beliefs of others (Echterhoff et al. 2009; Janis 1972).

Indeed, research shows that (sometimes) the function of beliefs might not be to accurately represent reality – but rather to facilitate traversing social distance by creating unity of minds. Studies have shown that individuals' desire to arrive at "shared reality" (Echterhoff et al. 2009) can overshadow the motivation to attain veridical beliefs about the world (e.g., Asch 1951; Turner 1991; for a meta-analysis, see Bond & Smith 1996). For example, when a leader announces that the sun is an almighty God, she could refer to a factual state of the world (in which case it would be relevant to consider evidence for and against this proposition); she may, however, implicitly mean something like "let us create a social group that would be united by the belief that the sun is an almighty god, leaving outside of our group anybody who does not believe so." The generation of unifying myths (Campbell 1991/1959), and more broadly, the generation of institutional facts (Searle 2010), have immense societal ramifications – they facilitate traversing social distance within groups of individuals and can thereby form the basis for creating large social structures such as tribes, religions, nations, and ideologies (which, in and of themselves likely contributed to the evolutionary fitness of humans; Henrich 2015).

Thus, despite the critical importance of prospection, the future is just one of many epistemic *and* social barriers that humans try to traverse. Importantly, similarly to Active Inference Theory, we argue that the process of traversing temporal distance shares many commonalities with the attempt to traverse other epistemic barriers. Therefore, in our subsequent discussion of prospection, we refer to this process under the more general heading of *mental travel*.

3.1. How does mental travel occur?

In the preceding section, we echoed theories of the predictive brain and suggested that the need to traverse the unknown is a central functionality of our brain and cognition. How is the feat of mental travel performed, and what is the role of our reservoir of abstract mental entities in this process? In the remainder of the manuscript, we present our answer to this question, which is mostly focused on deliberative acts of mental travel carried out by human beings.

As noted earlier, the act of abstraction creates a rich storehouse of various mental representations. Theories of prospection often stress that our memories of specific *episodes* are the critical reservoir of information upon which we build simulations of future worlds (e.g., Suddendorf & Corballis 2007). However, we wish to stress that all other types of abstract mental representations (e.g., *categories*, *predicators*, *hierarchies*, *scripts*) are just as important in mental travel. We shall refer to the reservoir of mental representations, which serve as the basis for the act of mental travel as the reservoir of *source representations*.

Importantly, when facing a specific problem of mental travel, people access their reservoir of source representations and generate a representation that models the specific problem at hand. For example, in order to decide whether to split the check in a restaurant on a first romantic date, we may access our knowledge regarding gender roles, the social background and the likely dispositions of the person in front of us, and recollect episodes of going on a first date. Based on this plethora of information we generate a target representation. In generating this representation, we make decisions regarding the relevant dimensions of the situation. For example, when deciding whether to split the check, I might consider the color of the clothes of my date to be an irrelevant dimension, but the gender and age of this person may seem to provide relevant information. Thus, the generation of the target representation constitutes an act of abstraction in and of itself.

However, generating a target representation is not the end of the story. Some mental processes must rely on this model of the situation and generate tenable predictions. A central tenet in many theories of prospection is the idea that people prospect by simulating future events (e.g., Barsalou 2009; Bechara et al. 2003; Gilbert & Wilson 2007; Redish 2013; Schacter et al. 2007). For example, when John is considering whether he should call or text his date from the day before, he could imagine calling her: "The phone is ringing, no one is picking up. I am ending the call. Now I don't know whether she's busy or whether she's

https://doi.org/10.1017/S0140525X19002000

not interested. This is very stressful." Having simulated this, John might decide that his best course of action is to send a text message on Facebook so that he could see whether the message was read.

The functionality of a simulation stems from the fact that the person running the simulation self-projects into it, that is, becomes an agent in the simulated situation. When the simulation is vivid and detailed (and thus similar to direct perception), reality-oriented processes (e.g., sensorimotor and affective/motivational processing, spatiotemporal associations, scripts) respond similarly to how they would react in real life (Gallese & Goldman 1998; Gordon 1986; Moulton & Kosslyn 2009). By "reading" the responses of the simulated self, one can generate new knowledge about the situation and decide how to act (e.g., Gallese & Goldman 1998).

Although the dating scenario described above represents a case wherein simulation is used to reason about consequences in a relatively novel scenario, simulation is also important in situations wherein organisms learn action-outcome contingencies from repeated experience (i.e., Reinforcement Learning; [RL]; e.g., Redish 2016; Skinner 1938; Tolman 1948). Research in this area distinguishes between "model-free" and "model-based" learning. In model-free learning, organisms choose their actions based on a representation that is simply an aggregation of previous hedonic outcomes associated with an action; as such, model-free behavior is impervious to sudden changes in the future (expected) utility of an action (for example, because of sudden devaluation of the rewards; Dickinson 1985). In contrast to model-free learning, many organisms can also learn to represent action-outcome contingencies in a "cognitive" (Tolman 1948), "model-based" (Daw & Dayan 2014) manner – namely, as a structured representation of the different possible states of the world, and the possible transitions between these states. Such target representations enable deliberative prediction processes (i.e., prospection), in which actions are selected based on consideration of future utilities rather than force of habit (Niv et al. 2006).

Much research suggests that model-based decision-making relies on simulation processes, or, as it is often referred to in the RL context – a process of "vicarious trial and error" (Redish 2016; Tolman 1948). During vicarious trial and error, organisms mentally test-out the different alternatives outlined in their model, and "read" the rewards and costs from the simulated scenario in order to choose the most favorable action. Indeed, single-cell recording studies have conclusively shown that when trying to choose between two arms in a T-maze, rats activate place cells that correspond to different routes they may take, thereby "mentally traveling" through the maze while sitting still (Amemiya & Redish 2016; Johnson & Redish 2007). Moreover, once the rat "arrives" at its goal, the simulation activates reward circuitry, allowing the rat to evaluate the action (van der Meer et al. 2010). Such findings provide compelling evidence for the importance of mental simulation in decision-making.

Despite the emphasis in the literature on the process of prospection via simulation (e.g., Barsalou 2009; Schacter et al. 2007), we contend that simulation is not the only route by which people can traverse the unknown; rather, people can also use theory-based inference; namely, "mentally travel" by relying on analogical reasoning (e.g., Gentner & Medina 1998; Hummel & Holyoak 1997; Reeves & Weisberg 1994) and on deduction (e.g., Evans 2003; Goodwin & Johnson-Laird 2013; Kuczaj & Daly 1979). For example, I can employ Sherlock-Holmes-like skills and reason that because my date studied at Vassar College, there is a good a chance that she has a negative view of traditional gender roles, which means that I should not offer to pay the check. Similarly, if I am to decide on whether to choose radio station X or Y, I do not necessarily need to simulate the expected outcomes. Instead, I can simply engage in proposition-based deduction (e.g., "station X plays classical music and station Y plays jazz; I am a person who prefers classical music; I should choose that which I prefer"). Unlike simulation, this form of inference does not require that one construct a representation that resembles sensory reality or experienced outcomes (e.g., hearing the music in my mind, feeling pleasure); therefore, *theory-based inference is more likely to rely on mental representations of higher abstractness such as highly abstract categories, intangible abstracta, and predicators*.

The distinction between theory-based inference and simulation has been a topic of much discussion within the literature on perspective-taking/mentalizing (i.e., the "theory-theory" vs. "simulation theory" debate; e.g., Apperly 2008; Gallese & Goldman 1998; Gilead et al. 2016; Gordon 1986), which, in our terminology, addressed the process of traversing social distance. Yet, it has been almost entirely absent from the literature on prospection.

When would people rely on theory-based inference and when will they simulate? As defined herein, simulation is a process that entails the projection of the self into a specific spatiotemporal context. Thus, whenever such a projection is difficult or unhelpful, we should predict that individuals will instead rely on theory-based inference:

When attempting to predict an event occurring at a specific spatiotemporal context (e.g., "where should we go for dinner?") simulation is a good idea ("I am eating this dish that I love. This is fun. Uh-oh, here comes the check.") However, when predicting outcomes that are not reducible to a representative scenario (e.g., "what will happen to divorce rates in the UK in the next five years?") simulation is largely irrelevant. One can simulate episodes contributing to a specific couple getting a divorce (e.g., an episode of infidelity) – yet such simulations do not really tell us what will happen on the national scale. Instead, it is better to use theory-based inference process such as proposition-based deduction (e.g., "the economy is in decline… economic hardship can increase marital distress").

When outcomes are not determined by human agents, the usefulness of being able to "read" the response of the simulated agent goes away, and theory-based inference should be more likely. For example, when trying to predict whether a specific greyhound will win at the races, individuals will likely perform some calculation based on past statistics and betting odds (rather than simulating the greyhound running, the wind blowing in his ears). Moreover, when outcomes are determined by members of a conflict group, these individuals are sometimes dehumanized to such an extent that they are deemed as no different from objects or animals (e.g., Haslam & Loughnan 2014); in such cases people will likewise often rely on abstract theory-based inferences (e.g., theories that explain all human behavior in terms of rewards and punishments) rather than on simulation.

When simulated others are dissimilar from the self, the utility of "reading" one's own responses to the simulated scenario diminishes, rendering the simulation less useful. For example, if I wish to figure out whether a close friend will be amused by a joke I can probably simulate my own response to the situation. In contrast, if I am interacting with a person who is socially distant (e.g., someone much younger) I might lack many of the specific parameters needed for the simulation and will be more likely to rely upon abstract knowledge and theory-based inference.

https://doi.org/10.1017/S0140525X19002000

When events have some precedence in one's reservoir of specific episodic memories (e.g., "how successful will I be as a comedian?") simulating these events can be useful (e.g., remembering times where your jokes fell flat). However, when no relevant episodes exist (e.g., "how successful will I be as a professional wrestler?") one can only rely on more abstract knowledge concerning the self (e.g., "I am an out of shape academic, this is not the typical demographic you see in professional wrestling; there might be a reason for that").

Despite these considerations, people often rely on simulation even when projection to a specific spatiotemporal context is less appropriate. For example, theory-based inference will probably be less likely when it requires the application of long, complex computations, and whenever cognitive resources are depleted (e.g., Stanovich & West 2000). Furthermore, because theory-based inference often depends on socially-acquired knowledge (e.g., heuristics and stereotypes, rules of normative probabilistic or logical inference, math, and so on), individuals who lack such knowledge (e.g., because of their young age, illiteracy) are more likely to use simulation during mental travel.

3.3. Interim summary: Section 3

In section 3 we proposed a bird's-eye account of mental travel. This account is somewhat at odds with several aspects of current theorizing: (i) contrary to some theories of prediction, we highlight that futurity is just one of the many epistemic barriers humans overcome (ii); contrary to theories that suggest that prediction is purely an epistemic process, we highlight that in humans, mental travel is often guided by the motivation to arrive at a state of shared belief with other humans – regardless of the truthfulness of those beliefs; (iii) contrary to theories that put a focus on episodic memory and simulation as the primary conduits of prospection, we highlight that an act of mental travel draws on a plethora of "source representations," some are relatively concrete and others are more abstract – and on both simulation *and* theory-based inference.

This account will now serve as the basis for our discussion of the role of different abstract mental representations in mental travel. In section 4, we explain why the diverse representational ontology we described in section 2 is indispensable in allowing mental travel. Finally, in section 5, we will review the empirical neuroscientific evidence that supports this account of representational diversity.

4. The members of our diverse representational ontology all help in meeting the challenges of mental travel

The challenge of mental travel stems from the fact that no target-representation will ever be identical to reality. However, what seems different on a concrete level could be seen as similar on a more abstract level. At the most rudimentary level, abstraction makes mental travel possible by introducing invariance among distinct experiences. Thus, although you cannot step twice into the same river, as the famous aphorism from Heraclitus goes, having a *multimodal abstractum* of water introduces stability into this endless variety, allowing the prediction that any time you put your feet in the river, it would feel wet. The organization of abstracta within *networks* allows us to predict which abstracta are likely to co-occur with a given abstractum (i.e., "this is a river; there must be fish around here.") The organization of categories within *hierarchical structures* allows us to draw inferences concerning properties of novel objects, based on their place in a hierarchy (i.e., "this is a fish; therefore, it must be edible like other fish."). The ability to organize *episodes* in temporal structures can allow us to re-play extended sequences and predict the conclusion of a possible course of action ("last time I ate fish, I ended up nauseous.") Finally, the cross-temporal organization of abstracta in *scripts* let us know what to expect and how to behave in situations that happen repeatedly ("whenever I feel nauseous, eating rice helps me feel better.")

In other words, the representational structures described earlier form the bridges that allow us to traverse uncertainty. In light of this, we believe that the link between abstraction and mental travel is fundamental to any consideration of these constructs; there is no mental travel without abstraction, and there is no need for abstraction but to support mental travel.

Beyond this fundamental claim, we argue that the different representational entities described earlier all play crucial roles in several (often-unrecognized) challenges associated with mental travel. We now turn to explicate this point.

4.1. The challenge of optimizing the accuracy/detail tradeoff of the target representation

In order for a target-representation to be functional, it must be accurate *and* detailed. When either condition is not met, the target representation is useless. For example, when trying to decide whether to go on a blind date, you may ponder, "what will my date look like?" If you tell yourself - "she will look like a human" you will be accurate, but this prediction will not provide you with any detail to guide your love life. In contrast, if you predict that - "she is 5'11 and has blonde hair," you have generated a detailed prediction, which could be useful except that it is less likely to correspond to reality.

As noted, the act of abstraction generates hierarchies of representations at varying levels of abstractness (e.g., "mammal," "human," "young human male with blonde hair"). This hierarchical organization of mental representations facilitates the construction of target representations in a manner that optimizes the accuracy-detail tradeoff.[27] By assessing the amount of knowledge at hand, one can tune the degree to which her prediction will be detailed and specific. Based on this logic, Construal Level Theory predicts that whenever people contemplate events that are more psychologically distant – and therefore involve more uncertainty – the optimal point of the accuracy/detail tradeoff moves towards higher-level abstractness and less detail (Shapira et al. 2012). Such logic is also consistent with the normative principle of Occam's razor,[28] according to which the best model is the one that introduces the least amount of (unsubstantiated) assumptions.[29]

Much research based on this theory shows that cognition abides by these normative principles and that the degree of psychological distance from an occurrence is a critical factor in determining the degree of representational abstractness. To give just a few examples, research has shown that people use more inclusive categories (Kruger et al. 2014; Liberman et al. 2002) and more abstract language (Bhatia & Walasek 2016; Fujita et al. 2006; Liberman & Trope 1998; Snefjella & Kuperman 2015) when imagining or making predictions of more distant situations (for reviews, see Liberman & Trope 2014; Trope & Liberman 2010).

Upon this view, the capacity to mentally travel to distant locations, planning farther into the future, imagining counterfactual worlds that are less similar to one's experiences, and taking the

https://doi.org/10.1017/S0140525X19002000

perspectives of more socially distal others – should co-occur with each other, as well as with an improved ability to form relatively more abstract mental representations. This co-occurrence should be evident across human evolution (e.g., the co-occurrence of greater interaction with distant others and the emergence of belief in imaginary moralizing deities; Norenzayan 2013), throughout ontogeny (e.g., the co-occurrence of delay of gratification and advanced reasoning ability; e.g., Rodriguez et al. 1989), as well as in covariance across individuals (e.g., co-morbidity between deficiencies in social perspective-taking and delay of gratification; e.g., Faja et al. 2013).

4.2. The challenge of making the construction of target-representations computationally efficient

Hoarding memories in a massive library of source representation will not do any good if transforming these into a target representation takes too long. What are the representational capacities that allow for efficient construction of target representations?

Consider the example of two international travelers. John goes to an airfare search website and seeks out a flight to Bangkok by looking at the entire list of flights, and considering whether each of them suits him in terms of price, time, and number of layovers. Jane is likewise searching for a flight to Bangkok but filters the flights and looks only at prices and times within the acceptable range. Alhtough both travelers will eventually find a flight, Jane used the taxonomic organization of the flight database and should find a flight more efficiently. Likewise, our database of mental representations is organized as a hierarchical taxonomy, allowing us to efficiently and quickly retrieve task-relevant source-representations from memory, and use them in order to construct target-representations. This provides the computational infrastructure that helps us to handle our massive storehouse of mental representations in an efficient manner (Bower et al. 1969; Cohen 2000).

If it is indeed the case that humans' capacity to organize abstracta within complex hierarchical structures increases retrieval efficiency, two straight-forward predictions follow: First, as demands for efficiency increase (for example, as individuals accrue more and more knowledge in a specific domain) people should be more likely to represent information in a structured, hierarchical manner (rather than, for example, strictly based on temporal order). Second, it should be the case that hierarchical organization in memory will indeed facilitate fast retrieval from long-term memory. The degree to which we organize our knowledge based on temporal contiguity or category-based hierarchies can be readily gauged by examining the clustering of items during free recall (Kahana 1996). However, memory research has not yet examined how the efficiency of subsequent retrieval is affected by the organization of material in long-term memory, nor has it examined the relation between the amount of information encoded (e.g., because of expertise) and the nature of its organization. These fundamental issues concerning memory volume, efficiency, and organization await further research.

Finally, like hierarchies of abstracta, *predicators* also allow efficient use of mental representations in that they contain free variables that make them applicable across various domains. For example, encoding "hang" as a predicator allows one to generate both "hang a picture" and "hang a towel" without requiring the learning of each of the specific instantiations. Indeed, research on primate language acquisition has shown that chimpanzees that were taught how to use signs as predicators (e.g., when they learned the meaning of a sign for "give me" as an entity that exists separately from "give me a banana" and "give me an apple") could process a multitude of different assertions (e.g., "give me juice," "give me carrots") – without the need be explicitly instructed on the meaning of each of these compositions, which would have been an intractable task (Savage-Rumbaugh et al. 1978).

4.3. The challenge of creating a richer repertoire of possible target-representations

If our repertoire of target representations would have been limited to previously-experienced events, mental travel would have been less useful. Theories of prospection stress that reshuffling of episodic memories allows us to generate many novel constructions, and thereby imagine previously unencountered events. According to our account, the reservoir of mental content we draw upon when constructing target-representations is not limited to episodic memories. Rather, humans use their full arsenal of source representations in order to enhance their ability to construe a multitude of alternative worlds. Below we outline three routes by which this is achieved.

4.3.1. Analogical transfer

One route by which a new target-representation can be generated is by importing some (but not all) aspects of existing source-representations into a new domain via analogical transfer (Gentner & Markman 1997). Consider the case of a newly-elected president, facing a crisis of increasing tension with a foreign country. Lacking experience in foreign policy, the president might seek an analogy to this political situation. She might invoke a playground *script* - "if you share your toys with other kids they will like you – I should offer concessions". Thinking that "foreign politics are like a playground" suggests that it involves a relation of reciprocity wherein you give X now, and you shall receive Y later – but does not imply that it involves a sandbox.

It is widely argued that without *relational predicators* that instantiate relatively abstract relations (e.g., "containment," "transitivity"; in our example - "reciprocity") the capacity for systematic analogical thought and inference would have been limited (see Gentner 1983; Penn et al. 2008) – and accordingly, our ability to construct novel target representations and traverse the unknown would have been less impressive.

4.3.2. Permutation

A second route by which abstraction broadens the scope of possible target-representations is by facilitating permutation (or "conceptual combination"; e.g., Fauconnier & Turner 2008). Consider the example of a cook contemplating a new dish. He can rely on an existing recipe and simulate changing the preparation method, ingredients, and so forth. We hope, one of the different combinations will yield a novel, tasty dish.

Again, it is his ability for systematic permutation of arguments within a *predicator* (e.g., fry(X); boil(Y)) that can greatly enhance the number of compositions he can create, and accordingly, the number of possible target-representations. Furthermore, the mere diversity in our representational capacities (i.e., the emergence of *intangible abstracta, multimodal abstracta*, *categories*) and the ensuing multitude of permutable objects very likely contributes to our capacity for representational generativity.

4.3.3. Cultural transfer

A third, critical route by which our diverse representational capacities increase the scope of possible target-representations is

https://doi.org/10.1017/S0140525X19002000

by facilitating language and cultural transfer (e.g., Cavalli-Sforza & Feldman 1981; Henrich 2015), namely, by subserving the processes of *symbolic interaction* (Mead 1934). Using language and other symbols (e.g., mathematical formula), we can efficiently adopt target-representations created by other minds. For example, my own experiences are unlikely to be helpful when generating a hypothesis regarding the outcome of a physics experiment; in such a case, I must rely on more abstract scripts that were gradually generated throughout human history, and on the cultural transfer of this knowledge.

It goes without saying that linguistic entities enable cultural transfer and that they are entities of a relatively high level of abstractness. In fact, language and abstract representations are so closely intertwined that they are sometimes thought of as synonymous. Language inevitably makes use of *categories* and *lemmas* that *symbolize* their referents. Furthermore, the generativity of language may rely upon the use of *predicators* (Bogdan 2008).

4.4. The challenge of decoupling the target-representation from the real world

In order for a target-representation to be functional, it must not be confused with reality. In the terminology of Nichols and Stich (2000), the segregation of the target-representation from the real world requires that the mental traveler creates a metaphorical "possible-worlds-box," the contents of which are distinct from reality. For example, when I imagine being chased by a tiger, or when I read about a person running away from a tiger, I should not confuse these thoughts with the presence of an actual tiger – as the appropriate reaction is quite different.

Theories of embodied cognition (e.g., Barsalou 2008) suggest that these different processes rely on the same representational bases, such that when we read about a tiger or imagine a tiger we activate the same perceptual and motor representations that become activated when we encounter a real tiger. Supporting this idea, much research shows that imagining visual stimuli activates the same neural regions involved in direct perception (e.g., Pearson et al. 2015; Redish 2013); furthermore, merely reading verbs that pertain to sensorimotor states (e.g., "kick," "toss") activates brain regions that control hand or leg movement (e.g., Hauk et al. 2004). Thus, it could have been easy to confuse simulation with reality, leading to maladaptive responses.

Non-psychotic adults do not typically confuse real experiences with imagined ones, although sometimes confusions do occur. For example, under some extreme conditions (e.g., sleep deprivation) normative individuals could be prone to hallucinations, which they deem to be reality (Babkoff et al. 1989). Furthermore, memory research (e.g., Johnson & Raye 1981; Goff & Roediger 1998) has shown that concrete, vivid simulations can easily give rise to false memories, wherein an imagined event is mistakenly thought of as one that really occurred.

There are several possibilities regarding the manner by which our brains distinguish between simulation and reality. One possibility is that the "possible worlds box" is implemented via different modes of processing in the *same neuronal populations*.[30] An additional possibility proposed by our model is that in order not to confuse imagination with reality individuals rely on representations of higher abstractness, supposedly subserved by *different neuronal populations* (Gilead et al. 2012; 2013). One advantage of abstract representations is their lower correlation with their referent. For example, the abstractum *animal* refers to many instances of animals; when one activates this abstractum, this activation may diffuse across many representations (e.g., cats, birds, and dogs), weakly activating each one of them. Thinking of *Danny's dog playing in the garden yesterday* activates a smaller set of representations, resulting in a more vivid mental image that could be more readily misinterpreted with truly seeing a dog.

If it is indeed the case that more concrete representations are more readily perceived as being real, it should be expected that people will make strategic use of this situation whenever they want to modulate the perceived factuality of displaced reference statements (i.e., statements that pertain to events that do not occur in the here and now). For example, it could be predicted that a comparison between the language used by prosecutors, who try to convince the judge and the jury that a transgression occurred, and defense attorneys, who have the opposite aim, will reveal that the latter use more abstract language (e.g., "Mr. Johnson lied about the value of the car he sold Mr. Smith" vs. "the defendant did not commit any wrongdoing in his business transaction with the plaintiff").

Finally, even when we accurately designate an event as being a simulation (e.g., something we heard about from others rather than experienced ourselves), we still need to designate whether this event is *fictional* or *non-fictional*. We regularly distinguish between biographies and a fairy-tales, documentaries, and non-documentaries, and (we hope) fake news and real news. It is likely that in order to make these distinctions we do not rely on different modes of neural processing or different neuronal populations; rather, we rely on the symbol-modifying capacities of predicators. For example, *intangible relational predicators* such as "not," "if," "imagine," and "believe" may play a critical role in keeping representations of fact, fiction, and hypotheticals logically and functionally distinct.

4.5. The specific challenges of traversing social distance

As noted, one of the most important types of mental travel is the traversing of social distance – the attempt to understand the beliefs, desires, and intentions of others, and to arrive at states of joint beliefs, desires, and intentions. In fact, several theories of cognitive evolution contend that the need to traverse social distance was *the* selection pressure that drove the development of the abstract representational capacities of humans (Deacon 1997; Dunbar & Dunbar 1998). We will now turn to explain how these representational capacities play a particularly important role in overcoming challenges associated with the attempt to traverse social distance.

4.5.1. Traversing the distance between two minds

Unlike any other dimensions of psychological distance, social distance can be traversed not only by using *theory-based inference* and/or *simulation*, but also by two straight-forward routes: (i) by simply asking the other person what is in her/his mind; (ii) by telling another person what is on your mind (thereby, making them to align themselves with the content of your mind). Clearly, representations that serve as the building blocks of symbolic interaction (i.e., *lemmas*, *categories, predicators*) are crucial for this type of communication-based alignment of minds (e.g., Austin 1975).

Furthermore, a unique complexity arises in the meeting of two minds, in that these minds engage in simultaneous, interactive predictions. In order to predict what my competitor in a chess match will do, I must try to represent what *she believes I believe* she will do, which depends on what *she believes I believe she*

https://doi.org/10.1017/S0140525X19002000

will do, and so forth (Camerer et al. 2002). The ability to process such complex recursion may rely on the iterative use of predicators that designate mental states, such as *believe* and *think*.

Indeed, the idea that mental state predicators are constitutive for performance on tasks that require mentalizing is supported by much research in developmental psychology (see Milligan et al. 2007, for a meta-analysis). For example, research has provided evidence for a causal relation between mothers' use of mental state verbs and children's subsequent performance on false-belief tasks (e.g., Ruffman et al. 2002).

4.5.2. Traversing the distance across an entire society

The ability to share mental states at larger scales (e.g., across nations, religions) provided the basis for large-scale cooperation in child rearing, agriculture, knowledge transfer, all of which contributed to humans' evolutionary success (Henrich 2015). Consider, for example, the emergence of a system of laws and norms – that undoubtedly facilitated social coordination. The use of abstract categories of behavior (i.e., murder) that are organized within a hierarchy of higher-order laws, constitutions, virtues and values (e.g., "sanctity of life"), allows potential offenders (and judges) to infer which behavior is prohibited –despite the infinite concrete ways by which a person can be murdered (see Hahn & Chater 1998). Moreover, the emergence of effective norms and laws may require impressive recursive capacities. For example, it has been argued (e.g., Searle 2010) that the ability to form a modern economy relies on intricate mentalizing: When a ruler declares that a note is valuable (i.e., a legal tender), citizens must believe that this ruler has the capacity to assign value to objects, that other people believe that it is indeed the case and that others believe that other people believe that is the case. As noted, this capacity may depend on the iterative use of *mental state predicators*.

Finally, in order for people to be willing to cooperate with non-kin individuals, it is useful to have credible displays of one's belonging and devotion to the group, a sort of "secret handshake" that allows entrance to an arbitrarily-established clique (Henrich 2015). The knowledge and practice of intricate myths and rituals (e.g., "thunder represents God's fury") are one especially potent means to achieve these goals (Campbell 1968). Importantly, such myths typically pertain to *intangible abstracta* (e.g., god) that cannot be deduced empirically from one's sense experience and logic – and therefore cannot be known without admission into the congregation and its specific teachings.

If it is indeed the case that beliefs pertaining to intangible abstracta are especially potent in generating strong bonds between strangers, it could be predicted that social groups that define their beliefs in terms of intangible ideas should become more cohesive. Future work could examine this prediction by investigating whether individuals who discuss their group membership in terms of more intangible ideas (e.g., "I am a Republican because I believe in *liberty*"), rather than in terms of concrete policy/action preferences (e.g., "I am a Republican because I don't want the government to take away my guns") exhibit greater solidarity with the group, and are more willing to engage in self-sacrificing behaviors in the name of the group.

4.6. Interim summary: Section 4

After describing an ontology of abstract representations in section 2, and describing the process of mental travel in section 3, in section 4 we have presented an account of how the diverse abstract representational entities that inhabit our mind all play a crucial role in mental travel. At this point we hope to have conveyed the message that mental travel likely relies on a plethora of diverse abstract representational entities – and that it may be insufficient to characterize the representational bases of mental travel by using relatively broad constructs such as "episodic memory," "cognitive model," or by relying on an undifferentiated, continuous hierarchy of mental representations of different levels of abstractness.

It is possible that alternative models of mental travel could account for the diverse competencies described herein, by assuming a more parsimonious representational toolkit (e.g., explaining the separation between reality and fiction without recourse to entities such as predicators). Such attempts could help refine or revise the model presented herein.

In the final part of this manuscript, we argue that if it is indeed the case that the process of mental travel builds upon these representational entities, understanding the neural bases of mental travel requires understanding the neural bases of these different types of mental representations.

5. Understanding the neural bases of the diverse representational architecture of the mind is essential to understanding the neural mechanisms of the predictive brain

Currently, functional neuroimaging is the central method by which cognitive neuroscientists study the neural bases of humans' predictive cognition. As implied by its name, functional imaging is most often employed to examine cognitive *functions* (e.g., predicting the future, memory retrieval, multisensory integration). Such functions are progressive mental acts that operate upon some object, a mental representation.[31] Thus, when we observe the predictive brain at work, we must remember to ask ourselves – does our observation reflect the working of a *type of cognitive function*, or the generation or use of a *representational type*? (see Wood & Grafman 2003 for a similar perspective).

In the final section, we apply our representation-focused, pluralistic perspective to the neuroscientific literature on predictive cognition. We begin by examining whether/how the neuroscientific literature coheres with our ontology of representational types, and then go on to demonstrate why a better understanding of the neural substrates of abstract mental representation may be crucial for research on the predictive brain.

5.1. Does neural evidence support the diverse representational ontology we have described?

In section 2, we provided an account of the different *types* of abstracta that correspond to different meanings ascribed to the term "abstract representation": *modality-specific abstracta*, *multimodal abstracta*, *categories*, several *complex structures,* and *predicators*. In this section, we will examine whether there is neuroscientific evidence that these entities are indeed distinct from each other.

5.1.1. Modality-specific abstracta

The existence of neural mechanisms specifically tuned to process modality-specific perceptual patterns has been demonstrated conclusively. Most notably, Hubel and Wiesel (1962; 1968) used single-cell recordings and showed that early visual processing operates in a hierarchical manner: neurons on the retina and the lateral geniculate nucleus respond to light at specific points

https://doi.org/10.1017/S0140525X19002000

of the physical world (Kuffler 1953); their projections converge to early visual cortex "simple-cells" that show selectivity to lines in a specific location and orientation; these "simple cells" converge to "complex-cells" which are orientation- but not location-specific (Hubel & Wiesel 1968).

This convergent architecture wherein neurons serve increasingly abstract modality-specific features is believed to continue until the generation of complex perceptual gestalts. Indeed, research has shown that along the inferior temporal cortex, cell assemblies converge to respond to gradually more invariant visual properties such as faces (e.g., Desimone et al. 1984) and entire scenes (e.g., Epstein & Kanwisher 1998). Despite these advances, we still do not have a complete computational account of how complex modality-specific objects, features, and relations (e.g., the image of a nose, the sound of a dog bark) are abstracted and stored. However, such an account may be inching closer, as scientists gain an increased theoretical understanding of the workings of deep (i.e., multi-layered) artificial neural networks (see LeCun et al. 2015).

A prevailing paradigm in neural network research is that experience-based representation learning occurs when a specific pattern of neuronal firing in a "lower-level" cell-assembly (i.e., pattern A) alters the strength of connections between the lower-level neurons and higher-order neurons (or a "deeper" layer) – in a way that increases the probability that a specific pattern in the deeper layer (i.e., pattern B) will recur in the future[32]; specifically, pattern B becomes more likely as activity at the lower level becomes more similar to pattern A. The higher-level layer contains fewer neurons[33] and therefore generates a more compact representation of the information in the lower-level layer. This means that different patterns in the lower layer are substitutable with each other in giving rise to the same pattern in the higher layer. For example, different percepts of triangles (e.g., equilateral, isosceles) which correspond to different patterns in the lower-level layer, may generate the same pattern of activity in the higher-level layer.

Some types of neural network architectures (especially those used in the PP theory; e.g., the Helmholtz Machine; Hinton et al. 1995) contain both bottom-up and top-down connections between the layers, such that activation of the higher-level layer (e.g., triangle) can re-generate the pattern of activity in the lower-level layer (e.g., an equilateral triangle). Endowing neural networks with such a "generative" capacity gives rise to many of the competencies observed in biological perception: Generative architectures allow the higher-level representation to predict future inputs, allow the network to "imagine" percepts, and retrieve modality-specific representations based on partial inputs (i.e., perform "pattern completion"; Hopfield 1982; O'Reilly & McClelland 1994). Such evidence suggests that the artificial neural network literature may indeed provide a good model of how modality-specific abstracta are generated and used during mental travel.

5.1.2. Multimodal abstracta

How are multimodal abstracta represented? According to the *modality-specific, widely distributed processing hypothesis* (e.g., Farah & McClelland 1991; Barsalou 1999; Kiefer & Pulvermüller 2012) which is inspired by the classic research into artificial neural networks (McClelland et al. 1986), multimodal representation are *not* subserved by specialized neural assemblies. Rather, the multimodal representation of, for example, "dog" is instantiated via reinstatement of patterns of activity across modality-specific cell assemblies in the visual, auditory, and somatosensory cortices. This hypothesis contradicts our model, in that it suggests that multimodal representations do not constitute a distinct representational type.

Despite the parsimony of this hypothesis, it is inconsistent with recent research. Whereas relatively posterior temporal regions subserve permanent (visual) representations, evidence from single-cell recordings performed on humans (e.g., Mormann et al. 2008; Quiroga et al. 2005; Quiroga et al. 2009) has shown that neurons further downstream in the Medial Temporal Lobe (MTL; i.e., in the hippocampus and in adjacent areas) selectively respond to *both* visual and verbal presentation of specific places and people, and thus may represent abstracted knowledge.

As noted earlier, whereas modality-specific abstracta can be grouped together based on perceptual pattern similarity, multimodal abstracta can be grouped together based on temporal contiguity. Much research has shown that the MTL, and especially the hippocampus, is involved in binding together perceptually distinct patterns based on the experience of their co-occurrence (e.g., Danker et al. 2016; Davachi 2006; Eichenbaum et al. 2007; Gottlieb et al. 2012; Sargolini et al. 2006). This research suggests that the hippocampal system may be critical in the encoding and retrieval of multimodal abstracta.[34]

Compelling evidence against the "modality-specific, widely distributed processing hypothesis" comes from extant fMRI research. In a comprehensive meta-analysis of functional neuroimaging studies of semantic processing, Binder et al. (2009) concluded that "semantic" knowledge lies within a wide, distributed network of regions, which includes the posterior inferior parietal lobe and the angular gyrus, middle temporal gyrus, the fusiform and parahippocampal gyri in the MTL, ventral, and dorsomedial prefrontal cortex, posterior cingulate gyrus, and the Left Inferior Frontal Gyrus (LIFG). These findings were replicated in research that examined patterns of neural activity that were common to the presentation of specific objects across different modalities (e.g., Fairhall & Caramazza 2013). The set of regions identified in these studies overlap with the set of brain regions referred to as the "Default-Mode Network" (DMN; Raichle et al. 2001). Importantly, with the exception of the fusiform and parahippocampal gyri (which are involved in the explicit mental imagery of concrete words; Gilead et al. 2013; Wang et al. 2010), this large swath of the cortex has no overlap with brain areas involved in sensory and motor processing – which suggests that multimodal abstracta are distinct from modality-specific abstracta (Binder & Desai 2011).

As noted above, the idea of layers of cell-assemblies that are abstracted away from their modality-specific instantiations is inconsistent with classic connectionist models (e.g., Farah & McClelland 1991). However, it is consistent with our view, as well as with more recent reincarnations of connectionist modeling (deep neural networks) that have shown that increasingly higher-level layers of cell assemblies can come to represent highly-abstract entities, that are coded in a (relatively) localized (rather than distributed) format (Bowers 2009) – and thus correspond to a-modal knowledge.

5.1.3. Categories

Multimodal objects that cannot be grouped based on spatial-temporal contiguity may nonetheless be deemed as substitutable, and generate *categories*. As noted, one does not need to encounter an image of a poodle with the associated word "mammal" to be able to categorize poodles as mammals (because poodles suckle

https://doi.org/10.1017/S0140525X19002000

milk and have hair). Categorical abstracta generate complex hierarchies of increasing abstractness, that often rely on the explicit linguistic transfer of socially-constructed criteria of substitutability (rather than on discovery via associative learning). However, is it the case that categories are neurally distinct from multimodal abstracta, as suggested by our model?

According to Rosch et al. (1976), people categorize objects into so-called "subordinate level" concepts (e.g., poodle), "basic-level" (e.g., dog) and "superordinate level" concepts (e.g., mammal). Unlike superordinate concepts such as mammal, basic-level concepts such as dog are more likely to be discovered by associative learning (e.g., observing various different dogs while hearing the word "dog") – and thus are more likely to evoke non-categorical representations (specifically, multimodal or modality-specific abstracta). Studies that have contrasted the processing of superordinate-level concepts such as mammal with basic-level concepts such as dog have found that processing the former involves greater activation within the Left Inferior Frontal Gyrus (LIFG) and the middle temporal gyrus (e.g., Raposo et al. 2012).

As noted earlier, categories also exist in the domain of goal-directed action. For example, the goal to "put on running shoes" is subordinate to the goal "going jogging," which is in turn subordinate to the goal of "maintaining health." Research into the functional architecture of the lateral prefrontal cortex (e.g., Badre & D'Esposito 2007; 2009; Badre et al. 2010) suggests that it is organized according to a hierarchy of abstractness wherein more anterior and inferior lateral frontal regions (e.g., LIFG) code more abstract, superordinate actions.

Furthermore, as noted, in contrast to modality-specific and multimodal abstracta, categories can refer to intangible entities. Therefore, another way to investigate whether categorical abstracta rely on a distinct neural population is by examining the neural substrates of processing intangible (vs. tangible) concepts. A comprehensive meta-analysis on intangible language processing (Wang et al. 2010) has shown that processing intangible words (e.g., justice, energy) as compared with processing of concrete words (dog, door) activates the LIFG and the anterior middle temporal gyrus.

Another way to investigate the neural substrates of categories is to observe brain activity as participants engage in tasks in which they attempt to find relations between stimuli that do not rely on spatiotemporal contiguity or perceptual similarity. Such acts are required during abstract analogical reasoning. Again, research into analogical reasoning points toward left-lateralized anterior frontal cortex as critical for this type of cognitive processing (e.g., Bassok et al. 2012; Bunge et al. 2004; Whitaker et al. 2018; see Hobieka et al. 2016 for meta-analysis).

Thus, research suggests that processing categories typically recruits anterior left-lateralized frontal and potentially frontotemporal cortical regions, implicated in linguistic processing/interaction (e.g., Kanwisher 2010). This provides tentative evidence that, indeed, it may be warranted to consider categories as functionally and anatomically distinct from multimodal abstracta.

As noted, most categorical abstracta (e.g., superordinate level concepts like "mammal," intangible concepts like "inflation rate") cannot be discovered without symbolic interaction with other people. The fact that the processing of categorical abstracta seems to rely on anterolateral frontal and temporal areas that are also involved in linguistic processing may reflect an affinity between the primary route by which criteria of substitutability are acquired and the neural systems upon which their associated abstracta eventually rely. Future research could investigate this hypothesis further by delineating the neural bases of acts of abstraction that rely on innate, personally experienced spatiotemporal contingencies, and socially-mediated criteria of substitutability – and examine how these three routes relate to subsequent retrieval of abstracta.

The acquisition of categorical knowledge via symbolic interaction has not been widely addressed in computational models of neural processing. The prevailing paradigm in neural network technology relies on inductive learning that associates between numerous instances (e.g., images of dogs and cats) and "labels" (e.g., the word dog or cat). Although this approach has led to impressive technological successes, the acquisition of categories in humans (and the intelligent behavior this affords) often relies on "one-shot learning" of criteria of substitutability, transferred from one mind to another. We do not present children with pictures of mammals alongside with the label "mammal" in order to teach them about mammals – we supply them a *definition* that allows them to recognize mammals and even imagine new instances of mammals. Thus, despite the impressive successes of technologies that followed the connectionist tradition, future advances in artificial intelligence research may require a rapprochement between "symbolic" and deep neural network architectures (see Lake et al. 2017 for a similar position).

5.1.4. Complex structures

As noted in section 2, theories of cognition have long suggested that representational entities bind together in an organized manner to form complex structures (i.e., temporal structures such as scripts and long-winding episodes, hierarchical taxonomies) that subserve mental travel. Given that our model views these structures as amalgams of abstracta, we probably should not expect to observe them as a distinct representational type – rather, as a product of functional connections between regions that subserve different types of abstracta described earlier, that is, as widely-distributed patterns of processing. Even so, according to our representational approach, investigators can still ask, for example, whether some of these complex structures predominantly rely on neural systems that subserve categories, multimodal abstracta, and modality-specific abstracta.

For example, one important question is how the neural representation of specific *episodes* differs from that of *scripts*. Whereas memory for particular episodes can contain information concerning specific perceptual details (e.g., the taste of the fish I ate), scripts represent information that is invariant across different particular situations (e.g., visits to different restaurants), and should rely less on modality-specific and multimodal abstracta. Indeed, research examining the neural representation of specific episodes versus general events (e.g., see Martinelli et al. 2013 for a meta-analysis) shows that the retrieval of specific episodes relies on visual areas (which subserve modality-specific abstracta) and on the MTL (which may subserve multimodal abstracta); in contrast, the retrieval of general event knowledge activates frontotemporal regions (which may subserve categorical abstracta). Importantly, contrary to some approaches in memory research, our perspective suggests that such findings should not be interpreted as representing a distinction between "semantic" and "episodic" autobiographical memory systems (e.g., Conway & Pleydell-Pearce 2000), but rather between systems that subserve categorical abstracta and multimodal/modality-specific abstracta.

Our pluralistic representational perspective can also inform attempts to reconcile findings concerning the neural substrates of semantic and episodic memory. Although facts ("semantic"

https://doi.org/10.1017/S0140525X19002000

memory) are often represented more abstractly than episodes, our model does not posit a one-to-one mapping between fact-knowledge and categories (or between event-knowledge and multimodal/modality-specific abstracta). For example, you can learn the whereabouts of the Empire State Building via symbolic interaction, when you are told that "the Empire State Building is in New York City"; in such a case, this fact will be encoded and represented by frontotemporal regions associated with categorical abstracta. However, the same fact can be discovered based on spatiotemporal contiguity experienced first-hand (e.g., repeatedly passing by the building when visiting NYC) or by repeatedly noting the unstructured association between the lemmas "New York" and "Empire State Building" in books and movies. Thus, our model suggests a *partial* dependence of semantic memory on neural substrates that are critical for episodic memory. In light of this, our model can explain why focal bilateral lesions to the MTL (critically associated with episodic memory) do not spare (nor do they obliterate) fact-knowledge, but rather cause partial anterograde and retrograde semantic amnesia (e.g., Lah & Miller 2008; Stark et al. 2005; Tulving et al. 1991).

Furthermore, our account predicts that research that attempts to localize the "semantic system" by asking participants to process words versus non-words (e.g., cloth vs. sworf) will activate the multimodal representations associated with lemmas, and therefore, should often evoke MTL activations (see, e.g., Montefinese 2019, for evidence supporting this prediction). As such, our perspective helps make sense of supposedly contradictory findings in the literature showing that the processing of "semantic" word meanings and "episodic" memory largely overlap (Binder et al. 2009).

5.1.5. Predicators

As noted, since Frege (1892/1952a), the existence of so-called "unsaturated entities," sometimes simply referred to as "concepts," has been posited by philosophers of mind and language – as these were thought to underlie the ability of mental representations to serve as functions that modify other mental representations. In contrast to the standard psychological approach (and consistent with approaches in linguistics; e.g., Kratzer & Heim 1998) our representational ontology posits that "unsaturated" entities (i.e., predicators) are distinct from "regular," "saturated" categories. Notably, this conjecture concerning the distinction between predicators and other categories has received very little attention in cognitive science (see Pylkkänen et al. 2011, for a discussion of this topic as an example for the disconnect between linguistic theory and neuroscience).

One way to examine the neural basis of predication is to look at the processing of verbs that differ in the number of arguments they require in order to be saturated (e.g., one-argument verbs such as "cringe," vs. two- and three-argument verbs such as "teach" that requires a specification of who taught who and whom). A recurring finding from such studies (e.g., Thompson et al. 2007; see Williams et al. 2017 for a discussion) is that the left angular gyrus increases in activity with increased demands for argument saturation – suggesting that this region may be crucial for predication. To the extent that the left angular gyrus indeed subserves predicators, our model predicts that it should be especially important in argument-manipulation processes such as those evident in logical deduction (e.g., application of predicators such as "if," and "or"), mathematical reasoning (e.g., the application of operations such as "minus" and "plus"), and ToM reasoning tasks (which may involve the application of mental state predicators such as "believe" and "think"). Indeed, in all these domains, there is evidence that lesions to the left angular gyrus result in significant behavioral decrements (e.g., Dehaene et al. 2003; Eimontaite et al. 2018; Zimmerer et al. 2019).

Thus, to summarize, the research reviewed herein provides compelling evidence for the distinction between modality-specific abstracta and multi-modal abstracta, and suggests a tentative model wherein the DMN, implicated in semantic cognition, may be parsed into (i) an MTL hub, that subserves multimodal abstracta; (ii) a left anterolateral frontotemporal hub, that subserves categorical abstracta; and (iii) a temporal-parietal hub, that subserves the unique class of "unsaturated" categories, namely, predicators. Much further work is needed in order to test, refine, or revise this neural model of mental representation and conceptual cognition; however, such an endeavor is essential in order to provide cognitive scientists with an accurate ontology of the representational entities that exist in our mind – and that subserve predictive cognition.

5.2. The neural bases of prediction

In this final section of the manuscript, we will apply our model to the neuroscientific research on prospection/self-projection, RL, and PP. In doing so, we highlight important issues that need to be addressed in future research.

5.2.1. Self-projection

Much research has investigated the neural bases of various types of goal-directed mental travel (often referred to as "self-projection"; Buckner & Carroll 2007) – such as contemplating future events (e.g., Addis et al. 2007), imagining hypothetical scenarios (e.g., Hassabis et al. 2007), and taking the perspective of others (e.g., Frith & Frith 1999). This research shows that it is appropriate to consider different types of mental travel as sharing an important common denominator. These various tasks all reveal neural activity localized to the medial prefrontal cortex, posterior cingulate cortex, and the angular gyrus – which are subcomponents of the DMN (for meta-analyses see Gilead et al. 2013; Spreng et al. 2009; for within-participants comparisons of mentalizing and prospection tasks see Spreng & Grady 2010; DuPre et al. 2016).

As noted, activity in these regions of the DMN is also associated with episodic memory retrieval (Kim 2016). In light of this, one of the most prominent theories of the DMN is that it is responsible for mental travel that occurs via simulations that rely on episodic memory (i.e., the "self-projection via episodic simulation" hypothesis; Buckner & Carroll 2007). Based on this influential hypothesis, some studies have interpreted activation of the DMN as a neural marker for the occurrence of episodic simulation processes (e.g., Peters & Büchel 2010; Tamir & Mitchell 2011; Tamir et al. 2015).

However, as noted, recent research suggests that virtually all of the different components of DMN also subserve the representation of *semantic* (rather than episodic) knowledge (e.g., Binder et al. 2009). In light of this, we suggest that the involvement of DMN regions in mental travel may be attributed to its role in subserving multimodal abstracta, categories and predicators. Whenever we see that the MTL regions of the DMN are involved in mental travel, this indeed may reflect the retrieval of particular episodes and a more vivid simulation process (Madore et al. 2016); however, when MTL activity is absent (and activity in the angular gyrus and anterolateral frontotemporal cortex is

https://doi.org/10.1017/S0140525X19002000

evident), we contend that that mental travel likely occurred via a "theory-based" inferential process, that relied on more abstract representations. Our account predicts that although episodic simulation can contribute to prospection, it is not essential. Indeed, this prediction is supported by research showing that individuals with extensive MTL damage have a preserved ability to reason about future events in a rational, normative manner (despite their deficiency in generating vivid simulations; e.g., De Luca et al. 2018; Kwan et al. 2012).

Construal Level Theory argues that because people are more ignorant about occurrences that are more psychologically distant, increased distance entails reliance on representations of higher abstractness, as these pertain to a greater number of possible alternatives and reduce error. Based on this, it could be predicted that when people contemplate more distant situations, they should rely on higher-level, superordinate categories which might be subserved by the LIFG and anterior temporal lobe. Partly supporting this prediction, Packer and Cunningham (2009) have shown that thinking of the more distant future resulted in activation in the LIFG and anterior temporal lobe. Similarly, Tamir and Mitchell (2010) and Majdandzic et al. (2016) have shown that activity in the LIFG increases as participants predict the beliefs of increasingly dissimilar others.

5.2.2. Reinforcement learning

As noted, it is widely held that there are two routes by which organisms can make decisions in the context of RL tasks: in habitual, "model-free" learning, the organism makes decisions using pre-computed values that were calculated based on the history of rewards associated with specific actions; in "model-based" learning, the organism predicts potential rewards by using a hierarchical mental representation that models the latent causal structure of events, and that allows deliberate prospection.

Much research provides compelling evidence that model-based RL indeed relies on simulation processes of the type discussed in the "self-projection" literature (e.g., Doll et al. 2015; Johnson & Redish 2007; Redish 2016; van der Meer et al. 2010). However, as we argued in section 3, episodic simulation may not be the only route by which humans prospect. With the help of representational conduits such as *categories, predicators,* and *scripts,* we can form innumerable different models (in our terminology, *target representations*) upon which different types of *theory-based inference* processes (as well as simulations) can operate.

Such theory-based inferences may also be important in repetitive value-based decisions (of the type discussed in the RL literature). Consider the example of a person who has to decide each morning whether to drive to work through the city (which is often busy with traffic), or the turnpike (which is less crowded, but entails a fee). As she enters her car, she can generate a vivid simulation of driving through the city (e.g., seeing the traffic slowly inching forward; feeling stressed by the prospect of being late). Such a simulation will likely rely on activity in regions associated with episodic memory retrieval (e.g., hippocampus) navigation and mental imagery (e.g., parahippocampal gyrus), as well as in regions involved in affective valuation (e.g., amygdala, anterior ventral striatum, orbitofrontal cortex). However, this person can also make her decision based on an abstract, theory-based inference – that does not require her to generate a facsimile of reality (e.g., "this is summer so more people are on vacation, which means that the city might be less crowded; however, rent prices keep soaring, so people must have less disposable income to go on vacation; the city will be swamped.") Such inference likely relies on regions that subserve the retrieval and systematic manipulation of categorical abstracta (i.e., left-lateralized frontotemporal regions). Thus, although research has conclusively shown that humans do rely on concrete simulation during model-based RL – this does not rule out the possibility that theory-based inferences may also support model-based decisions. The extent to which humans rely on such abstract inference processes in the context of value-based decisions remains an open question.

A fuller understanding of the abstract representational bases of cognition (and their diverse, hierarchical nature) is also crucial for future research on model-*free* learning (which do not rely on deliberation and prospection). RL studies have identified that what animals learn is dependent on the learning situation (e.g., responses that were acquired in a specific situation do not necessarily transfer to other situations). In light of this, as noted by Redish et al. (2007, p. 790), a major question faced by the decision-maker is: "not a decision-process question – Should I act or not? – but rather a cognitive question – Which situation am I in? … the recognition that the agent's current situation shares properties with previous (similar) situations."

Our perspective suggests that the same exact environmental situation can be categorized/construed at different levels of abstractness. For example, a couple attending a basketball game can construe the situation as "a night out in town" (which may entail a craving for a nice glass of red wine) or "sitting on a stadium seat at Madison Square Garden" (which would evoke craving for a hot dog). Research within Construal Level Theory has demonstrated that increasing psychological distance from a situation makes people construe it more abstractly. Much behavioral work within this framework has highlighted how this regularity in the process of construal/situation-recognition can explain various behavioral outcomes in the domain of decision-making (e.g., variation in intertemporal discounting – Liberman & Trope 1998; Trope & Liberman 2000; self-control failures – Fujita et al. 2006; melioration – Pick-Alony et al. 2014; exploration–exploitation decisions – Halamish & Liberman 2017; Yudkin et al. 2019).

The fact that the same situation can be categorized at different levels of abstractness means that it is important to distinguish between situation-recognition processes that are primarily based on modality-specific or multimodal properties (e.g., a basketball game is deemed similar to other crowded, loud gatherings) and categorical characteristics (e.g., a basketball game is deemed similar to other "competitive situations"). Our model predicts that the former will rely on retrieval of representations in the hippocampal network and sensory cortices, and the latter will rely on the left lateral frontotemporal cortex. Future research is warranted in order to examine these predictions.

Such considerations also give rise to novel, potentially important research questions. For example, we know that "model-free," habitual behaviors can be triggered by situation-recognition processes that pertain to modality-specific abstracta (e.g., a red light in a Skinner box). However, can habitual responding be likewise activated by recognition processes that rely on categorical abstracta? For example, can model-*free* RL mechanisms be used to condition a person to light a cigarette whenever she (specifically) reads the works of French existential philosophers (but not of Russian Existentialists)?

5.2.3. Neurobiological models of predictive processing

We argued that abstract representations have a causal role in cognitive processing. Moreover, echoing the "symbol-processing

https://doi.org/10.1017/S0140525X19002000

view" (Newell & Simon 1972), we suggested that some abstract mental representations (predicators) have the capacity to interact with other representations in a systematic manner, thereby giving rise to a sort of "language of thought." Like language, such a system may be best modeled by its own unique principles (i.e., rather than being modeled by the same principles that explain visual perception and motor action).

The historic alternative to this view is the connectionist (or "subsymbolic") perspective (McClelland et al. 1986). Connectionist models adopt a parsimonious architecture, wherein a simple set of computations suffice to explain all aspects of cognitive functioning – from basic perception to language. On this view, the unique characteristics of different types of abstract mental representations, and the unique functions they may afford, are epiphenomenal to understanding the algorithms of the mind. In recent years, developments of connectionist models (i.e., deep neural networks) have demonstrated the utility of this approach, by achieving remarkable success in solving real-world computational challenges.

Neurobiological process-models of PP have been inspired by neural network models (e.g., Hinton et al. 1995; Hinton et al. 2006) and likewise often adopted rather parsimonious mechanisms. Most notably, similar to connectionist models, a major strength of PP theory (and specifically, Active Inference Theory) is that it shows how the complexity of cognition can naturally arise from a canonical computation repeated across different layers of a single continuum of representational abstractness.

Contrary to accounts that seek a "neat" organization of cognition, we advanced the case of the so-called "scruffies" (see Schank & Abelson, in Clark 2013; Marcus 2009), namely, those who believe that cognition has a "varied bag of tricks" (Clark 2013). We do not deny the importance of subsymbolic processes and representations but rather endorse a pluralistic perspective according to which both subsymbolic and symbolic representations play crucial roles in cognition (see Griffiths et al. 2010 for a similar pluralistic approach). Specifically, we argued that when we zoom-in into different layers in the hierarchy of mental representations, we reveal meaningful neurobiological distinctions (e.g., neural substrates) and functional distinctions (e.g., different roles in prediction) – whose explication and integration with theories of the predictive brain could further develop theory.

One area where such potential integration could be especially meaningful is in the domain of mental health. PP models of psychopathology have provided compelling accounts of mental illnesses as stemming from aberrant belief-updating dynamics *between* different layers of the representational hierarchy. Specifically, the theory points at dysfunctionalities in the weight given to update signals between higher- and lower-level layers, as giving rise to "false beliefs," and phenomena such as psychosis and depression (e.g., Adams et al. 2013; Clark et al. 2018). However, it is possible that some pathologies of belief formation can also be attributed to aberrations of the unique computations that occur *within* specific layers.

For example, if categories and predicators give rise to a "language of thought," one could ask how the dynamics within such a highly-abstract system are related to illness. As illustrated by Asimov's (1950) depiction of the robot Herbie, who went mad once he realized that he could not abide by some of his imperatives without breaking others, empirical research suggests that irreconcilable inconsistencies within one's system of abstract beliefs about the self (i.e., "cognitive dissonance"; Festinger 1962) can lead to emotional distress. When cast in terms of Bayesian belief dynamics, confidently believing two contradictory abstract propositions about the self (or "splitting"; Kernberg 1975; e.g., "I am a bad person," "I am a saint") may generate instability in the top-down predictions of the self-evidencing architecture of the brain, giving rise to psychopathology (e.g., unstable self-worth; borderline personality disorder). As a response, more resilient individuals may construct a novel model that accounts for conflicting beliefs (e.g., "I am human, and humans are multi-dimensional"); once such a model is selected, it may "suppress" prediction errors and increase the stability of the system.

Importantly, one unique principle of the layer of language-like mental representations is that it is directly accessible via symbolic interaction (e.g., through self-talk, or via talking with others). Indeed, since Freud it has been argued that "speech therapy," in which patients are given novel interpretations of their experiences can alleviate distress. Such symbolically-mediated belief-updating processes – that are posited to be fundamental to many aspects of human life – are currently not addressed within the PP literature. As such, we suggest that attempts to model the dynamics that occur within humans' system of symbolic representations present a future challenge for PP models, and a critical test of their ability to provide a comprehensive account of psychological distress.

6. Concluding remarks

In recent years, cognitive scientists have increasingly taken to investigate the role of prediction as a fundamental process of cognition. It is widely held that humans' advanced capacity for prospective thought is subserved by similarly advanced capacities for abstract mental representation. In light of this, in the current manuscript, we attempted to provide an explicit and integrative account of the abstract representational bases which allow humans to "transcend the here and now."

Based on theorizing and research in philosophy and in the neural and cognitive sciences, in sections 1–2 we provided a bird's-eye view of the ontogeny and ontology of abstract mental representations. In line with influential theories in the predictive brain framework, we suggested that abstractions are built as a hierarchy, ranging from the highly concrete to the highly abstract. However, importantly, we argued that the different computational challenges associated with prediction give rise to important qualitative differences between different types of abstractions that exist along this hierarchy, generating meaningful diversity in the representational substrates of the mind.

Specifically, echoing views from philosophy, we suggested that the representational hierarchy can be parsed into three qualitatively distinct levels: *modality-specific* representations that are primarily instantiated on perceptual pattern similarity; *multimodal* representations that are primarily instantiated on spatiotemporal association; and *categorical* representations, that are primarily instantiated on social interaction. These representational elements serve as the building blocks for more *complex structure*s discussed in cognitive psychology (i.e., episodes and scripts, networks and hierarchies), and are the inputs for mental representations that behave like functions, and have been discussed mainly in linguistics, namely, *predicators*. We offer this ontology as our best attempt at an "elemental table" of the mind – to be revised, extended, or replaced.

We provided two types of arguments to support our model: first, we explained how the different elements in this ontology are all needed in order to account for humans' impressive predictive cognition; this argument may be contrasted by alternative

https://doi.org/10.1017/S0140525X19002000

theoretical models that explain how the capacity for mental travel can be explained by a different, perhaps simpler, ontology. Second, we examined how the neuroscientific evidence coheres with our account and highlighted future research that could further support or contrast our model.

If our conceptualization is somewhat right, then this means that theories of the predictive cognition have endorsed an overly simplistic picture of the representational architecture of the mind and that this simplicity may hinder the ability of these models to account for behavioral and neural phenomena. We highlighted several directions by which our diverse representational ontology could guide future research into the predictive brain (e.g., regarding the functionality of the DMN, the role of theory-based inference in RL, and the role of symbolic representations in PP models of psychiatric illness).

In conclusion, the evolving framework of the predictive brain offers an opportunity for greater integration across the cognitive sciences. Psychologists, neuroscientists, and philosophers have long been working on piecing together ideas concerning the representational bases of cognition. The current manuscript attempts to build a bridge between this rich history and the newly evolving framework of predictive cognition. It is our hope that this bridge will assist scientists in their future mental travels.

Notes

1 And several other species (e.g., Clayton et al. 2007).

2 Although episodic memory is more detailed and concrete than semantic memory, it is nonetheless *declarative*; namely, information that can be readily put into words, and as such (per our discussion of abstraction later on) may be considered more abstract than procedural/non-declarative memory.

3 Per definition, there are potentially innumerable mental states active at any given moment; these states can extend for milliseconds or a lifetime.

4 This analysis of mental states in terms of their logical relation with the mind and the world comes from Searle's (e.g., 1979; 1983) discussion of *intentional states*.

5 Clearly, some mental states (in their general form; e.g., wanting food) exist prior to acts of abstraction performed throughout an individual's life; they stem from biological evolution.

6 Our use of the term belief/desire/intention is inclusive, refers to logical relations between the mind and the world, and does not entail specific claims about the representational apparatus subserving beliefs or about their human uniqueness (following Dennett & Haugeland 1987).

7 Satisfaction is the brief moment where our desires are met, or when novel information evidences our beliefs. When put in the formal terms of active inference theory (Friston 2005), satisfaction of an intentional state can be thought of as minimization of free energy by the realignment of the sensory units of the so-called "Markov blanket" (Kirchhoff et al. 2018) through the altering of internal belief states, or through acting on the world.

8 The definition is inclusive in that it does not require consciousness or awareness of the formation of this belief.

9 In Bayesian epistemology, abstraction can perhaps be seen as the formation of a belief that different causes (e.g., chocolate, ice-cream) of an internal state can be reduced to a single cause, a more parsimonious theory – a process akin to model selection and dimensionality reduction.

10 See Griffiths et al. 2010, for a discussion of the merits of competence-level analysis of cognition.

11 The notion of concreta is related to Frege's (1892/1952b) notion of *Bedeutung* (referent).

12 It is important to distinguish between *internal* criteria of substitutability, which are the rules as they are represented in the mind that performed the act of abstraction (i.e., things that are made of chocolate are tasty); and *external* criteria of substitutability, which are the actual rules that determine the criterion of substitutability (e.g., unbeknownst to the agent, the reason objects A and B are considered by her as tasty is that they are sugary).

13 In the terminology of Medin et al. (1993), *respects for similarity.*

14 According to our definition, abstraction is fundamentally a situated, context-dependent process (e.g., Barsalou 1983; Hintzman 1986; Smith & Semin 2007). There is no "correct" hierarchy of abstracta, or some pre-defined criterion for selecting the set of possible objects that will satisfy a belief or desire. For example, in a specific context, "cake" may be deemed substitutable with "ice cream" and both would form the concreta of "dessert"; in another context, "cake" may be deemed substitutable with "bread," and both will form the concreta of "things that you bake." In other words, the abstractum is, in principle, a unit that is defined by an ad-hoc use. This, however, does not preclude the emergence of regularly-used abstracta between individuals or between different points in time, nor does it preclude the emergence of abstracta that are more central than others (see later the discussion on the emergence of complex representational structures).

15 Whereas influential models (e.g., Tversky 1977) have been long able to address the problem of similarity-based categorization, attempts to precisely model theory-based categorization (Murphy & Medin 1985) have been notoriously challenging (see Pothos & Chater 2002). Approaches that highlight the need to provide the simplest theory (i.e., "the simplicity principle"; Chater & Vitanyi 2003) provide a promising avenue for research on this problem (Pothos & Chater 2002). Future application of the PP framework to higher-order cognition may be able to provide formal accounts of theory-selection processes during theory-based categorization.

16 In the terminology of Deacon (1997), which follows in the footsteps of semiotic theory (Peirce 1931), such abstracta exhibit *Iconic* reference. The ability to perform such an act of abstraction is also referred to as the "Høffding step" (1892).

17 In the semiotic terminology such abstracta exhibit *Indexical* reference.

18 In the terminology of the animal learning literature, episodes are types of stimulus-stimulus associations (see Holland 2008 for a review).

19 In the semiotic terminology such abstracta exhibit *Symbolic* reference.

20 Note, however, that predicators are presumed mental (rather than linguistic) entities, and their existence may be independent of language acquisition (see Carruthers 2002, for discussion).

21 This view, which echoes perspectives from linguistics, diverges from the prominent view in psychological research that assumes that predication (or "conceptual combination") entails a symmetrical interaction between two concepts of similar standing.

22 Some version of feature-based exemplar or prototype theories may nonetheless be able to support predication (see Prinz 2012a, for a discussion).

23 Such a capacity may also subserve the ability to generate "ad-hoc categories" (see Barsalou 1983).

24 Formally, the discrepancy between our representations of the world before and after we interact with it.

25 "Free-energy minimization" (Friston 2005) is an information-theoretic description of the behavior of every self-organizing biological system (see Kirchhoff et al. 2018 for further discussion); when applied to predictive processing accounts of the brain, it describes the process of the long-run minimization of prediction errors.

26 Such priors are supposedly predominantly derived from processes that occur at the evolutionary timescale.

27 In the terminology of Goldsmith et al. 2002, the tradeoff between accuracy and grain size; see also Rosch et al. 1976.

28 This logic is also consistent with Active Inference Theory, that elaborates how free-energy minimization entails minimizing the difference between the accuracy of a model and its complexity, and thus, entails a preference for minimally complex models.

29 In the terminology of Active Inference Theory, such an account might mean that with increasing psychological distance, the precision assigned to descending signals at lower-levels of the hierarchy is attenuated relatively to those at a higher level.

30 For example, it is suggested (Pezzulo et al. 2017) that within the hippocampus, shifting from a mode of neuronal activity that generates theta oscillations to a mode that involves "sharp waves and ripples" (e.g., Ylinen et al. 1995), may reflect a shift from reality-oriented processing to simulation.

31 It should be noted that such a distinction between function and representation is disputed within sub-symbolic architectures (e.g., McClelland et al. 2010).

https://doi.org/10.1017/S0140525X19002000

32 An idea derived from research into long-term potentiation of neuronal synapses; e.g., Kandel 2001; Lømo 1966.
33 More accurately, it is to the very least a "sparser" architecture (see Field 1994).
34 It could still be argued that the hippocampus simply provides a "transport hub" between different modality-specific representations subserved by *modality-specific* cortical systems (e.g., McClelland et al. 1995).

Acknowledgments. We wish to thank Greg Murphy, Gary Marcus, Dedre Gentner, Tal Eyal, Ran Hassin, Yoav Bar-Anan, Ruvi Dar, Yakir Levine, Jochen Weber, Rani Moran, Alexa Hubbard, Raphael Gerraty, Dave Kalkstein, Miri Perkas, Zvee Gilead, Talya Sadeh, Britt Hadar, Hadar Ram, David Kashtan, Chelsea Helion and the anonymous reviewers for immensely helpful feedback.

Open Peer Commentary

Other and other waters in the river: Autism and the futility of prediction

Matthew K. Belmonte[a,b]

[a]The Com DEALL Trust, Bangalore 560043, India and [b]Division of Psychology, Nottingham Trent University, Nottingham NG1 4FQ, UK.
belmonte@mit.edu http://www.mit.edu/~belmonte/

doi:10.1017/S0140525X19003194, e122

Abstract

Autism has been described as a neural deficit in prediction, people with autism manifest low perceptual construal and are impaired at traversing psychological distances, and Gilead et al.'s hierarchy from iconic to multimodal to fully abstract, socially communicated representations is exactly the hierarchy of representational impairment in autism, making autism a natural behavioural and neurophysiological test case for the prediction–abstraction relationship.

Gilead et al. lament that theories of abstract cognition have been left unintegrated in part because of a lack of terms of discourse common across branches of the cognitive sciences, or even between social and biological aspects of psychology. There is indeed some irony in this all too lowly construed approach to the cognitive science of construal and abstraction, distinct threads of which have been appearing in the history of cognitive science for at least the past seven decades. Our story begins with Witkin's (Witkin & Asch 1948; Witkin et al. 1962) notion of *field dependence* in perception and psychophysics, and its subsequent relationship to gestalt-orientated cognition and to social affiliation and perspective-taking (Witkin & Goodenough 1977). This same idea of a concrete–abstract representational axis cutting across perceptual and social aspects of cognition was recapitulated by Frith (1989) and Frith and Happé (1994) as *central coherence* in describing both autism's decontextualised detail-orientated perceptual stance and its likewise decontextualised egocentric social perspective. Around the same time the idea was introduced to social psychology by Trope (1989) first as an account of dispositional trait versus situational state explanations of others' behaviour, then extended to effects of temporal and other psychological distances on what Trope et al. had come to call perceptual *construal* (Trope & Liberman 2003), the term adopted in the rest of this commentary.

The syndrome of autism, along with its dimensional extension to individual differences in autistic (or what Witkin called field-independent) traits, exemplifies this association between construal and psychological distance: Spatial, temporal, social, and hypothetical distances resurface as autistic differences in mapping between allocentric and egocentric space (Conson et al. 2015; Frith & de Vignemont 2005; Hamilton et al. 2009; Pearson et al. 2014; Ring et al. 2018), impulsivity and executive disinhibition (Hill 2004), social perspective-taking and other aspects of cognitive empathy (Baron-Cohen 1995), and repetitive-behavioural aversion to unpredictability and change (Gomot & Wicker 2012). Gilead et al. relate the distinction between raw perceptual observations and elaborated cognitive models (abstracta) to the contrast between detail-orientated, first-person *simulation* and abstract, allocentric *theory* in predicting the behaviour of the world; impairment in prediction when constraints are underspecified, dynamic, or real-time – as is the case in social cognition – has been identified time (Courchesne & Allen 1997) and again (Sinha et al. 2014; Van de Cruys et al. 2014) as a unifying feature of autism which may drive the co-occurrence of anxiety and rituals, perceptual dysmodulation, visuomotor deficits, slowed orienting of attention, and undifferentiated processing of stimuli regardless of task-relevance. Because autistic predictions tend to be founded more on iconic, concrete perceptual data rather than on abstracta, they evoke many violations of expectation in instances where observations would match the broad strokes of an abstract model yet fail to match these minutiae (Van de Cruys et al. 2014). This hyper-reliance on iconic representations produces a style of cognitive inference by *bricolage*, that is, by effortful construction and maintenance of complex representations and ideas bottom-up from the underlying details and instances (Belmonte 2008a), which are preserved *in lieu* of abstracta (Belmonte 2008b). This flattening of Gilead et al.'s hierarchy of abstracta implements a cognitive style adroit at recognising relationships amongst numerous, low-construal percepts, described by Baron-Cohen et al. (2009) as "systemising." Although it can confer superiority at detail-orientated disciplines, this systemising style imposes such a great cognitive representational load that it cannot scale. Because predictions based on inappropriately detailed cognitive models frequently evoke mismatches with observations, and such errors of accidental detail are not differentiated from errors of essence (Van de Cruys et al. 2014), the world amounts to a constant chaos of Heraclitean flow in which one's expectations are always and inexplicably wrong, sabotaging social and other domains of reward and thus impairing learning and development. It's no surprise, then, that Gilead et al.'s hierarchy of representational qualities – from concrete, iconic, modality-specific impressions, through multimodal convergences (Brandwein et al. 2013; 2015; Ostrolenk et al. 2019), to socially communicated, categorical abstractions (Beker et al. 2018; Feldman et al. 2018; Smith et al. 2017; Stevenson et al. 2017) – is exactly the hierarchy of perceptual and representational abnormality in autism.

All this evidence shows Gilead et al.'s ontology of abstraction and prediction to be consistent with historical concepts and findings, and with what we know about autism, its prime test case. But retrospection is the game of Monday-morning quarterbacks – what of prospective predictions, and experiments yet to be performed? Drawing together all these strands can relate behavioural and neural aspects of prediction and abstraction, psychological distance and construal, with corollary implications for cultural and sex differences in cognition: Gilead et al. speculatively peg the default-mode network as the home of their cognitive abstracta, although the true locus may lie rather in this network's interactions with other control networks. The default-mode network is constitutively active in autism (Kennedy et al. 2006), perhaps reflecting constant and largely fruitless attempts at predictive modelling (Raichle 2015) of accidental detail, associated with low-construal impulsive action (Shannon et al. 2011) and anxious affect (Simpson et al. 2001a; 2001b).

The female advantage in default-mode network deactivation in reward contexts (Dumais et al. 2018) seems consistent with autism's association with male-typical cognition (Baron-Cohen et al. 2005), linking construal to cognitive sex differences. And Witkin (1979) himself noted that construal variations can be a function of culture; indeed individualistic cultures are associated with a more systemising bias (Markus & Kitayama 1991; Nisbett & Masuda 2003) and collectivistic cultures with higher construal (Boduroglu et al. 2009; Masuda & Nisbett 2006). One might predict, then, associations of individual trait construal level (a.k.a. autistic traits, field dependence), situational state construal level, sex and/or gender, and individualistic/collectivistic culture with the frequency and/or duration of dynamic coupling of default-mode with attentional and executive control networks (Ryali et al. 2016). The degree of network coupling would reflect individual and situational differences in the bias and range of model-driven feedback versus environmentally bound feedforward cognitive control of perception, action, and affect, and would be measurable with fMRI, or perhaps EEG/MEG (Kitzbichler et al. 2015). Such a study would afford an opportunity to reconstrue (as it were!) as a neurophysiological variable the diversity with which individual humans walk the tightrope between Aristotelian category and Heraclitean instance, between Lacan's (1966) *le symbole* and *la chose*.

Touch me if you can: The intangible but grounded nature of abstract concepts

Anna M. Borghi[a,b] and Luca Tummolini[b]

[a]Department of Dynamic and Clinical Psychology, Sapienza University of Rome, Via degli Apuli 1, 00185 Rome, Italy and [b]Institute of Cognitive Sciences and Technologies, Italian National Research Council, Via San Martino della Battaglia 44, 00185 Rome, Italy.
anna.borghi@uniroma1.it luca.tummolini@istc.cnr.it

doi:10.1017/S0140525X19003091, e123

Abstract

Thinking about what the senses cannot grasp is one of the hallmarks of human cognition. We argue that "intangible abstracta" are represented differently from other products of abstraction, that goal-derived categorization supports their learning, and that they are grounded also in internalized linguistic and social interaction. We conclude by suggesting different ways in which abstractness contributes to cement group cohesion.

In their thought-provoking article, Gilead and colleagues provide a much needed unified conceptualization of abstraction and show how modality-specific, categorical representations up to scripts and more complex mental structures form a hierarchy of increasingly abstract mental representations. Recently, we have proposed an embodied/grounded theory of "abstract concepts," the Words As social Tools (WAT), that similarly highlights the role of linguistic and social interaction (Borghi & Binkofski 2014; Borghi et al. 2017; 2018a; 2018b; 2019a; 2019b).

Here, we focus on one of the highest rungs in their abstraction ladder – categorical representations – and address three related issues: the distinction between "tangible" and "intangible" abstracta, the role that goal-derived categories might play in the ontogeny of intangible abstracta, and the role of language and sociality for their representation.

First, we argue that, even if the level of conceptual abstractness is extremely variable and context dependent, intangible abstracta differ from other tangible categorical representations, like superordinate concepts, more than the authors acknowledge. Intangible abstracta are considered among the hallmarks of human cognition, indeed many authors argued that concrete (tangible) and abstract (intangible) concepts rely on at least partially separate systems (Shallice & Cooper 2013). It is to highlight the difference between intangible abstracta and other categorical representations that we have distinguished between *abstraction* (the process leading to categorizing and representing concepts hierarchically, e.g., "collie"–"dog"–"animal") and *abstractness* (the process leading to the formation of abstract concepts like "freedom" whose referents are more detached from sensory modalities and are not bounded single entities) (Borghi et al. 2019a). In our view, sensorimotor, interoceptive, linguistic, and social dimensions are relevant for both intangible and tangible abstracta, but to a different extent. Linguistic and social experience is certainly relevant also for tangible abstracta: we learn from others that bats are mammals. Still, "mammals" refer to a collection of perceivable exemplars that ultimately ground their meaning; the case is different for concepts like "freedom": Linguistic and social inputs become essential to cohere otherwise disparate perceptual experiences. Thus, the suggestion that, compared to multimodal abstracta, "categories" engage areas linked to linguistic processing as the left inferior frontal gyrus (LIFG) fits well with our view, but holds more for intangible abstracta than for superordinate concepts (Borghi et al. 2019a; see also Dove 2018).

We believe, however, that this evidence does not support the claim that their representation should be considered as amodal. Much evidence shows that the processing of intangible abstracta also activates sensorimotor areas (Sakreida et al. 2013). Relying uniquely on meta-analyses that focus on areas activated to a greater extent by abstract than by concrete concepts risks to downplay the importance of areas engaged by both; furthermore, meta-analyses typically collapse across different kinds of intangible abstracta, and this can exclude important sensorimotor information. Finally, neuroimaging studies generally include highly concrete items and collapse intermediate and highly abstract concepts, thereby increasing the variability for abstract concepts (Pollock 2018). Evidence shows that specific abstract concepts are grounded in event-based, interoceptive, introspective, sensorimotor areas: for example, temporal concepts rely on perisylvian locations generally identified in time perception studies (Lai & Desai 2016), spatial processing areas in the posterior parietal cortex are activated for numerical concepts whereas emotion concepts engage regions of amygdala and orbitofrontal cortex involved in emotional experience (Desai et al. 2018). Finally, claiming that their representation is amodal is especially problematic for tangible abstracta like superordinate concepts. Behavioural studies have indeed shown that superordinates activate multiple exemplars and their sensorimotor features through an instantiation principle (Borghi et al. 2005; Heit & Barsalou 1996; Murphy & Wisniewski 1989).

Second, we believe that *ad hoc* and goal-derived categories (Barsalou 1983; 1985) should play a more important role within the framework outlined by the authors. We propose that the capability to form and use goal-derived categories constitutes one of the bases enabling the formation of intangible abstracta. Learning intangible abstracta implies the ability to form and acquire categories that do not have single objects as referent, whose members are not perceptually similar. Such flexibility is present in goal-derived categories, which generally cross the boundaries of standard taxonomic categories, as in the case of "birthday presents" that may include exemplars like flowers, animals, and artefacts. Learning goal-derived categories can provide a bootstrapping mechanism useful for further acquisition of intangible abstracta.

Third, several studies conducted in our lab and in other labs have shown that abstract concepts activate linguistic experience through the involvement of the mouth motor system. Participants rate abstract concepts, particularly mental state ones, as more associated with mouth than hand actions (Ghio et al. 2013; Granito et al. 2015); consistently, fMRI has shown that the mouth motor system is engaged during processing of these concepts (Dreyer & Pulvermüller 2018). When participants process abstract concepts, they are facilitated in responding using the mouth rather than the hand (e.g., Borghi & Zarcone 2016; Mazzuca et al. 2018). Furthermore, blocking the mouth selectively influences abstract concepts acquisition, as demonstrated by studies with children who used the pacifier until late (Barca et al. 2017; in press). We have proposed that mouth involvement might be linked to a mechanism of social metacognition (Borghi et al. 2018b; 2019a; Fini & Borghi 2019; Villani et al. 2019): Because abstract concepts are more complex, we feel less competent (Shea 2018) and need help from authoritative others (Prinz 2012b), preparing ourselves to ask for information. Gilead et al. propose that abstractness has an important evolutionary function because groups that define their beliefs in terms of intangible ideas might become more cohesive. We argue that our social metacognition mechanism can help to increase group cohesion. Differently from multimodal abstracta, the intrinsic complexity of intangible abstracta induces individuals to assess their own competence and to rely more on others. This mechanism can be powerful in creating social bonds, because it helps individuals to recognize the role of others as dispensers of knowledge, and induces competent others to share their knowledge. It can also contribute to explain social hierarchies, based on the different competences people have in matters relevant for their group.

On the implications of object permanence: Microhistorical insights from Piaget's new theory

Jeremy Trevelyan Burman

Theory and History of Psychology Department, Heymans Institute for Psychological Research, Faculty of Behavioural and Social Science, University of Groningen, Groningen 9718NG, Netherlands.
J.T.Burman@RUG.NL https://www.rug.nl/staff/j.t.burman/

doi:10.1017/S0140525X19002954, e124

Abstract

The authors' arguments reflect the dominant traditions of American Psychology. In doing so, however, they miss relevant insights omitted during the original importation (translation and popularization) of the foreign sources that informed the theories they built upon. Of particular relevance here are Piaget's last studies. These are presented to unpack the meaning of "object permanence" as a kind of representation.

"Microhistory" is a method used by historians: we investigate something small to derive new insights that reveal something big (see Burman 2017, pp. 119–120). Here, I focus on the target article's mention of Piaget's object permanence as part of its engagement with representation. It is an aside, there, and not used to advance the authors' argument. But it could have. As a result, we can use the microhistorical method to fill in some of the theoretical details that would otherwise be missing from subsequent discussions.

First, though, some basics: object permanence is the result of constructing the logical operation of *identity*, such that sensations at time "n" following a particular motor interaction become associated with the same sensations that are observed at time "$n + 1$" following the same interaction. The consequence is then the abstraction – note my slightly different usage – of a representation (an "object") that persists over time. Piaget (1977/2001) ultimately called this process "empirical abstraction" (*passim*).

Once objects have been abstracted (constructed) from movements-and-sensations (phenomena), they become the new basis for the child's conception of reality. The sensorimotor is thus replaced with the concrete. Abstraction then changes too. From the persistence of represented-objects are abstracted sets, and laws, and these can in turn be applied to *imaginary objects* (themselves also a kind of representation). The resulting *reflected abstractions* and *meta-reflections* are also treated *as if* they were real. Therefore, the world changes again.

Several scholarly commentaries explain the details (see, especially, Campbell 2001; 2009; Moessinger & Poulin-Dubois 1981). Here, though, I want to focus on the big picture; to use the micro to exemplify the macro. And even though the insights I cite are from the end of Piaget's life, quite a lot happened in those final years. They need to be considered together.

An important related observation is that what Piaget was doing in Geneva is not identical to how his work was understood and popularized by American Psychologists. As a result, it has become common to refer to the divergence as "Piaget's new theory" (following, especially, Beilin 1992). This involved several changes, made in parallel, but it is typically characterized in the secondary literature as involving a shift from logics of extension to those of intension (Davidson 1988; 1993; Ducret 1988; after Piaget & Garcia 1987/1991).

For us, this change enables the treatment of abstraction as involving functional identities (implication, signification, and meaning) rather than strict identities (between sensations or objects in themselves). And that was in turn made possible by the replacement of stages, at the start of the new theory period, with levels of relative incompleteness.

This is the so-called neo-Gödelian turn in Piaget's theorizing: It replaced the popular staircase metaphor of cognitive development with "an upwardly broadening spiral" (translated by Burman 2016, p. 762). It also clarified the notion of abstraction by enabling the recognition of identities across levels in that spiral.

The easiest way to understand the part of this that matters for our purposes is to read it through Bruner's (1960) reinterpretation of Piaget for American teachers. In particular, I am thinking of the "spiral curriculum" (pp. 13, 52–54) that became so influential during the post-Sputnik period of education reform.

In a spiral curriculum, the same topic is revisited at different levels of complexity across different grade levels. New insights are then derived by reflecting on the similarities: Although the externally-provided educational structures are different in different grades, the functional consequences for their understanding of the issue-at-hand are similar. Hence, the levels in this spiral are comparable by virtue of their reference to the identity of the pedagogical object being considered.

Something like this occurs during cognitive development too, in Piaget's new theory, except that the scaffolding is provided endogenously: *functionally*-identical consequences are derived from quite different interactions, treated across levels, such that the lineage of related representations is unified by different kinds of abstractions. (In Piaget's later language, this is possible because the comparisons involve "morphisms" [see, especially, Piaget et al. 1990/1992].) This in turn enables the construction of correspondences between different functional-structures and then generalizations within, between, and across levels (Piaget 1980b; Piaget & Henriques 1978). Therefore, non-overlapping areas can be filled-in. And that is why Piaget's (1980a) conception of dialectics includes periods of calm between its dialectical punctuations; how you get the appearance of discontinuous stages despite continuous change.

This is part what's missing in the authors' view of representation, but which we can see as a result of adopting a microhistorical approach. That also affords the main historical criticism of such work: contemporary authors are too embedded in the post-Sputnik popularization of Piaget as a theorist of *cognition*, and insufficiently grounded in what the Genevans were actually doing. As a result, they miss the same things that were omitted during Piaget's original importation into American Psychology: the neglected "foreign invisibles" (Burman 2015).

In other words, the article is missing those aspects of abstraction that Piaget used throughout his interdisciplinary program. Yet, this is equally as important as his psychology (Ratcliff & Burman 2017). The recognition of what's necessary – given the intensional interpretation of identity as functional equivalence – drives the exploration of what's possible so that what's constructed is *new* rather than a *copy* (Piaget 1981/1987; 1983/1987). This thinking also informed the basis for his misunderstood evolutionary-developmental theory, updating the Baldwin effect (Piaget 1976/1979; discussed by Burman 2013; 2019). And the same formal principles afforded new comparisons between child development and scientific change (Piaget & Garcia 1983/1989).

Thus, for Piaget, abstraction and representation were not only properties of *cognitive* development. Nor were they solely a function of the development of knowledge in general. Rather, they were a function of *life* (Burman, in press). And so research like the authors' contributes to more than just psychological knowledge. It also advances the role of psychological theorists at the frontier of the extended evolutionary synthesis; the move toward "evo-devo/psych-know" (Burman 2019, pp. 292, 307). Or rather, they can do so long as we are able to recognize the unexamined implications of their sources.

Language as a mental travel guide

Charles P. Davis[a,b], Gerry T. M. Altmann[a,b] and Eiling Yee[a,b]

[a]Department of Psychological Sciences, University of Connecticut, Storrs, CT 06269 and [b]Connecticut Institute for the Brain and Cognitive Sciences, University of Connecticut, Storrs, CT 06269.
charles.davis@uconn.edu http://charlespdavis.com
gerry.altmann@uconn.edu http://altmann.lab.uconn.edu
eiling.yee@uconn.edu http://yeelab.uconn.edu

doi:10.1017/S0140525X19003182, e125

Abstract

Gilead et al.'s approach to human cognition places abstraction and prediction at the heart of "mental travel" under a "representational diversity" perspective that embraces foundational concepts in cognitive science. But, it gives insufficient credit to the possibility that the process of abstraction produces a gradient, and underestimates the importance of a highly influential domain in predictive cognition: language, and related, the emergence of experientially based structure through time.

Transcending the present moment – referred to by Gilead et al. as "mental travel" – is indisputably central to human thought: It encompasses not only predicting the future, but also traversing distance on several other psychological dimensions. In order to predict, we need to abstract, and in order to abstract, Gilead et al. argue, we rely on a diverse toolkit comprising three distinct levels of representation. We are skeptical that there exist qualitatively distinct levels of a representational hierarchy, and instead suggest a graded continuum from "modality-specific" to "categorical" representations. Further, we contend that a key factor in promoting development of this gradient – underappreciated in the proposed toolkit – is language.

As Gilead et al. suggest, a consequence of the drive to reduce prediction error is the emergence of representation at multiple levels of abstraction. But, these levels need not be qualitatively distinct: for example, evidence suggests that conceptual knowledge is represented on a posterior-to-anterior gradient along the temporal lobe, with modality-specific information becoming less salient more anteriorly (e.g., beagle–dog–animal; for a review, see Davis & Yee 2019). Critically, the role *language* plays in processes of abstraction and prediction deserves greater recognition (for discussion, see Yee 2019). Language is perhaps *the* quintessential example from human cognitive behavior of (levels of) abstraction, prediction, and the relationship between them. Both language comprehension and production may build on more general predictive mechanisms involved in action planning and understanding (e.g., Pickering & Garrod 2013; see also Altmann & Ekves 2019), and language is, by definition, abstracted away from objects and events. And in addition to providing a useful model of prediction and abstraction at multiple levels of representation, language plays a *functional role* in facilitating these functions, and thus, "mental travel."

Many formal models of abstraction in language exist, but here we focus on work describing prediction and the emergence of abstract category structure as a function of accumulating knowledge of the contexts in which experience is grounded (Elman 1990; see also Altmann 1997). Jeff Elman's work with the simple recurrent network (SRN) is the quintessential example from a *computational* standpoint of abstraction, prediction, and the relationship between the two (Elman 1990; 1993). Through accumulated experience of sequences of words, categorical distinctions such as between parts of speech (e.g., noun and verb), and between classes of nouns and verbs (e.g., edible objects and intransitive verbs) *emerge* in a network given the task of predicting the next word in the sequence.

Gilead et al. perceive an insufficiency in models exhibiting an "undifferentiated, continuous hierarchy of mental representations of different levels of abstractness." Yet Elman's SRN *was* undifferentiated *computationally* (hidden layer units all functioned identically). After learning though, it was not undifferentiated *functionally*. Similarity relationships in its equivalent of the external world (language input to the SRN) were maintained in its acquired internal representations, and these allowed the SRN to predict the space of possible inputs at the next point in time. Hierarchy was only categorical to the extent that hierarchical clustering is categorical (different clusters would exhibit different hierarchies). Abstraction in Elman's work was *graded*, meaning generalization was graded also – a desirable property in a probabilistic world. Importantly, and unlike Gilead et al.'s framework, which have since been shown, in deep recurrent neural networks, are general principles of *learning and development* (Elman et al. 1996).

The emergence of increasingly abstract representations (not just in language) may rely on domain-general neurobiological mechanisms for tracking systematicities across space and time (for discussion of how one such mechanism may apply to abstract concepts, see Davis et al. 2020). However, a problem for any experience-based model of abstraction is how we sample enough of the world to track those systematicities and converge on shared meaning. Here, language comes in again: It allows us to experience more of the world than we could *via* direct experience alone. Experiencing spoken, signed, and written words – and their distributional patterns of co-occurrence both with other words in sentences and with the real world – opens a window into other people's (embodied) experiences. Distributional language statistics are a rich source of knowledge (e.g., Louwerse 2008), enabling us to make predictions about things not directly experienced.

Language also facilitates prediction and abstraction in ways non-linguistic thought does not. For example, labels may penetrate through the representational gradient by operating directly on mental states (Elman 2009). Although classical thinking holds that language is merely a means to communicating our thoughts, more recent work has shown that language has a functional role not only in higher-order thought, but also perception (for a review, see Lupyan 2012). A consequence of language's influence across the gradient of abstraction is that concepts do not operate only at the modality-specific level: labels may (among other things) help integrate modality specific information in higher-order association areas. Gilead et al. cite meta-analytic findings that lexical-semantic tasks tend to activate higher-order brain regions far removed from modality-specific areas (Binder et al. 2009) as "compelling evidence" against distributed, modality-specific models of cognition. But, these activated higher-order regions are integral to multimodal integration and conceptual access *via* labels. Furthermore, because there is diversity in the modalities in which different things are experienced (e.g., sunsets visually, vs. thunder auditorily), conceptual representations reflect that diversity (e.g., Davis et al. in press). Thus, when experiments average over dozens of diverse concepts, activity in the various modalities that contribute to each one is likely to be washed out.

Abstraction is a process, and this process engenders a gradient, not qualitatively distinct levels in a representational hierarchy. Moreover, an account emphasizing "representational diversity" to address how humans use prediction and abstraction to transcend the present moment should recognize the ubiquitous role of language. Not only does the scientific study of language processing offer well-tested, formalized frameworks for understanding how abstract structure emerges (e.g., Elman 1990; see also Altmann 2017), but language itself plays a functional role in facilitating "mental travel" *via* its integral role in prediction and abstraction.

Representation, abstraction, and simple-minded sophisticates

Peter Dayan

Max-Planck-Gesellschaft, Max Planck-Ring 8, 72076 Tübingen, Germany.
dayan@tue.mpg.de https://www.kyb.tuebingen.mpg.de/publication-search/60427?person=persons217460

doi:10.1017/S0140525X19002942, e126

Abstract

Bayesian decision theory provides a simple formal elucidation of some of the ways that representation and representational abstraction are involved with, and exploit, both prediction and its rather distant cousin, predictive coding. Both model-free and model-based methods are involved.

Bayesian decision theorists (BDTs), a group which active inferencers might beneficially pupate to join, are sophisticated simpletons. The simpleton half of this oxymoron comes from the straightforward inferential crank that they turn to generate behaviour (Berger 1985): agents should characterize their probabilistic beliefs about the state of the world; evaluate the expected present worth of the potential long run future consequences of their available choices or actions given this characterization; and make an appropriate choice in light of these evaluations. BDTs are sophisticated because done correctly, this leads to optimal behaviour in both individual and collective (Harsanyi 1967) settings, and because of the statistical and computational complexities they have to overcome to execute each of these steps correctly.

From a formal perspective, we can see the centrality of predictions about the future (a rider that will seem less odd shortly) – because it is the portended worth of those consequences that matter. Indeed, agents' very characterizations of the current state of the world should only make distinctions that make a difference in terms of what the future might hold (Dayan 1993; Littman et al. 2001). However, BDTs only need to predict evaluations – predicting in more detail what will happen is at most a means to this particular end.

I hope that Bayesian decision theory helps put all the rich representational and process distinctions in the target article into slightly starker light. In terms of representation, abstraction is a useful, and indeed sometimes normative, approach to the complexities mentioned above. Throwing away distinctions that do not matter (or perhaps do not matter very much) allows one to generalize predictions about future worth (perhaps approximately), obviating more learning, more computation, or indeed both. One might quibble about the particular forms of abstraction considered here – for instance, the article frequently flirts with deterministic, rather than probabilistic, criteria for substitutability. This would seem likely to be somewhat too rigid in most circumstances.

Second, in terms of processes, we can see that neither simulation theory nor the "theory–theory" that the article puts in partial competition with it, are really fundamental constructs – because we only really need to predict evaluations rather than actual future outcomes. It is this observation that underlies the sorts of model-free reinforcement learning (RL) to which the target article refers (Sutton & Barto 1998), and which can also exploit rich representations. Of course, there are statistical benefits (though computational costs; Daw et al. 2005; Keramati et al. 2011; 2016; Pezzulo et al. 2013) to model-based (MB) RL – in which more elaborate aspects of the future are predicted as a means of making long-run evaluations. However, one might note that even this conventional sort of MB RL already includes the sort of flexible incorporation of inferential abstraction which is referenced – there is nothing that requires any vividness of simulation. Perhaps, the term "cognitive model" in MB RL might have seemed a bit overly ascetic. Equally, one might note the active investigation of how episodic and semantic contributions to various forms of RL are integrated (Collins & Frank 2012; Gershman & Daw 2017; Lengyel & Dayan 2007).

A third elucidation concerns the fact that predictive coding models (MacKay 1956; Rao & Ballard 1999) also consider prediction about the present – a sort of ersatz prediction that should be kept conceptually completely separate from predictions about the future. That is, such models specify hierarchical abstractions of the current state as a way of analysing that state. They do this by considering how this state might have been generated, that is, how it might have been synthesized. An example of this sort of analysis by synthesis (Neisser 1967) is to consider performing computer vision to analyse a visual scene into its underlying contents by determining all the settings of the graphics engine in a computer game that could synthesize the scene. Each setting would provide a description of the objects, their positions, the lighting, the location of the observer, the shot noise, etc., that could have produced the scene. One way to perform this analysis is to start from some likely settings, predict what the scene should look like if those settings were indeed responsible, look at how the actual scene differs (this is the prediction error), and change the settings accordingly. Ultimately, though, it is the analysis that matters (a conclusion that the target article steps around somewhat balletically).

Analysis by synthesis turns out to be a powerful idea about how to create abstractions in what is known as an unsupervised manner (Hinton & Sejnowski 1999). Furthermore, the target article points to aspects of such generative models that could usefully be structurally far more sophisticated (although the sorts of probabilistic programming notions that are becoming popular in some circles; Goodman et al. (2012) arguably generalize even the highest order representational construct considered by the target article, namely predication). However, even the earliest thoughts about unsupervised learning (Marr 1970) were suffused with concern about the fundamental lack of justification for these sorts of representational ideas for the task of making good decisions for the future – a dilemma that is, however, not resolved here.

In sum, I applaud the authors for their lucid challenge to overly simplistic notions of representations and processes. Abstraction is of tremendous benefit in many ways to real prediction and thus real control, and therefore much work in RL is attempting to find ways of determining and exploiting appropriate representational structures both within single domains of decision-making, and across multiple such domains.

Abstracting abstraction in development and cognitive ability

Andreas Demetriou

Cyprus Academy of Sciences, Letters, and Arts, University of Nicosia, Cyprus, 1700 Nicosia, Cyprus.
ademetriou@ucy.ac.cy

doi:10.1017/S0140525X19002930, e127

Abstract

We focus on the theory of abstraction proposed by the target article. We suggest that abstraction varies at different levels of learning, cognitive development, or cognitive ability. We argue that this theory does not specify how abstraction is done at each of these levels. Because of these weaknesses, the theory cannot explicate how individuals differ in mental time travel at different phases of life or different levels of cognitive ability.

1. Introduction

Launched by Aristotle (Back 2014), interest in abstraction resurged in psychology (Burgoon et al. 2013; Reed 2016) and spread to artificial intelligence (Saitta & Zucker 2013) because it is important for understanding and problem-solving in variable environments. The target article proposes a theory of abstraction as a basis for mental time travel. In this theory, representations are parsed into three categories (*modality-specific*, *multimodal*, and *categorical representations*, emerging from perceptual similarity, spatiotemporal contiguity, and social interactions, respectively). These are building blocks, abstracta, for the construction of *complex structures* (episodes, scripts, and hierarchies) predicated by language at increasingly higher levels of abstraction. The structures are sources of predictive processing. Substitutivity, the recognition that two or more elements may stand for the same referent, is the fundamental mechanism of abstraction, underlying generalization, reduction, categorization, and analogical induction. Other theories stress other mechanisms as the basis of abstraction, such as identifying invariant central characteristics (Burgoon et al. 2013) and discrimination (Reed 2016). We argue that the proposed theory of abstraction (sect. 1 and 2) is weak in several respects, thereby failing to account for predictive cognition (sect. 3 and 4).

2. Abstraction in learning and development

Abstraction is partly indeterminate: its very operation changes its subsequent state and products, rendering future abstractions different. Therefore, any theory of abstraction must account for how learning, in the short-term, and development, in the long-term, change abstraction. This theory does not involve any such provisions: it does not explicate how modality-specific abstractions are formed at the first place nor does it account for their integration into multimodal and categorical representations. Simply naming the origins of substitutivity is not enough. We need to know how innate abstractions emerge out of interactions with the physical properties of the world at the first place, how they are redefined by personal experience, and how they are reshaped by social interactions. We also need to know how and when abstraction processes change.

The brain evidence invoked as supportive (sect. 5) is as global as the abstraction theory itself. Localizing different forms of abstraction in different brain regions, even if accurate, says nothing about abstraction itself. We need to know how brain regions operate and speak to each other when forming abstractions. Optimum connectivity defines the precision of abstraction (Raju, in preparation); brain rhythms may be the lexicon and their coordination the syntax of brain language (Buzsaki 2010; Demetriou & Spanoudis 2018).

Cognitive development is the development of abstraction. Thus, it was central in all cognitive developmental theories. In Piaget's (2001) theory, abstraction is the engine of equilibration, the central mechanism of cognitive development. In our theory, abstraction is part of a tripartite system involving, additionally, alignment processes generating relations feeding abstraction, and cognizance, awareness of mental processes and their products, allowing metarepresentation yielding abstracta (Demetriou et al. 2018a). Levels of cognitive development reflect the ontogeny of abstraction. The proposed theory must explicate how modality-specific, multimodal, and categorical abstractions emerge at successive cognitive developmental cycles. In infancy, before language, modality-specific abstractions generate the primary material that in toddlerhood, with language, will be weaved into complex multimodal realistic representations. Later, in primary school, these multimodal representations are organized into rule-based categories increasingly predicated by language. In adolescence, rules and predications are meta-represented by principles predicating truth, protecting from deception (Demetriou & Spanoudis 2018).

Bayesian inference underlying abstraction dominates early in learning or development (Tenenbaum et al. 2011); logical mechanisms in analogical and deductive reasoning dominate later (Demetriou & Spanoudis 2018). Their precise proportion and overall representational profile involving *modality-specific*, *multimodal*, and *categorical representations* at different phases of learning and development is not specified in the target article. Also, it is important to specify how reflection integrates modality-specific abstractions into multimodal abstractions and how awareness of abstraction processes underlies categorical predication. Learning research shows that guided and reflected upon relational processing generates abstraction in different domains (Jee & Anggoro 2019; Papageorgiou et al. 2016).

3. Abstraction in individual differences

In classical theory of intelligence, individual differences in intelligence reflect differences in abstraction. The very notion of general (Jensen 1998; Spearman 1904) or fluid intelligence (Cattell & Horn 1978) is basically abstraction coming under different names (e.g., Spearman's eduction of relations and correlates). Individual differences in intelligence reflect differences in how far individuals progressed along the developmental course of abstraction outlined above (Demetriou & Spanoudis 2018). Therefore, higher intelligence reflects more increasingly flexible, to-the-point, abstractions employed for long-term predictions about the world.

Processing and representational efficiency constrain abstraction. For instance, attention guides abstraction to relevant information and working memory provides the field where it occurs (Demetriou & Spanoudis 2018; Demetriou et al. 2018b). Attention and working memory lapsuses may misdirect or disorganize abstraction. The paper is again silent about how abstraction interacts with these processes.

4. Mental time travel

In conclusion, the theory presented in the target article does not specify how abstraction is done at different levels of development, ability, or learning, what is the difference in abstraction between different levels of intelligence, and how we learn to abstract in different contexts or contents. Thus, the theory is weak in explicating mental time travel in relation to different types of representations. For instance, the episodic representations of the pre-language infant allow some time of prospection: Infants have predictive models of their behavior *vis-à-vis* their environment that protect them from falling or colliding with objects; however, they may dangerously err in unfamiliar environments. The realistic mental representations of the toddler allow social prospection: toddlers predict others' behavior based on their knowledge of others' mental states; however, they may seriously err if others' behavior is based on different values about these mental states, which they do not know. Primary school children use abstraction to foresee their daily activities. Adolescents build prospective models of their life as adults. The target article is silent about how different types of representations and related abstractions engender different types of perspective in mental time travel at different phases of life.

Is it always so? Unexpected visions

Jan B. Deręgowski and Benjamin W. Tatler

School of Psychology, University of Aberdeen, Aberdeen AB24 3FX, UK.
j.b.deregowski@abdn.ac.uk
b.w.tatler@abdn.ac.uk

doi:10.1017/S0140525X19003170, e128

Abstract

If we consider perceptions as arising from predictive processes, we must consider the manner in which the underlying expectations are formed and how they are applied to the sensory data. We provide examples of cases where expectations give rise to unexpected and unlikely perceptions of the world. These examples may help define bounds for the notion that perceptual hypotheses are direct derivatives of experience and are used to furnish sensible interpretations of sensory data.

This is a *rider* provoked by the meritorious paper. It concerns the central issue – the origin and nature of the hypotheses triggering the predictive processing. Gregory (2009) in his studies of illusions suggested that previous experience influences such hypotheses. Segall et al. (1966) entertained this notion and tested proneness to illusions of cultural groups differing in experience of carpenteredness and openness of the environment, an approach followed by other cross-cultural researchers (Deręgowski 2017). There are, however, well known instances when the origin of the hypotheses does not quite match such paradigm. That is, rather than illusions illustrating the role of underlying hypotheses on producing the percept based upon expectations furnished by past experience, there are cases where the percept that arises from the illusion directly contradicts that to which we have become accustomed – even to the extent that we accept as our percept a form that violates known rules of perception. For example, the Necker cube fluctuates in depth, although the main depth cue it presents is that of oblique segments which can be seen as either receding to the left or to the right on the background of the paper, which furnishes no definite depth cues. Removal of this inert background does not remove the tendency to hypothesise, as the following observations concerning Zagloba's puzzle described below show.

A funnel looked into monocularly is seen for what it is: its converging walls receding towards the spout. After a short time, such percept changes spontaneously – perceptual inversion occurs and the funnel is no longer seen as a funnel, but as a tepee whose outer walls are the inner surface of the funnel and whose smoke outlet at the top is the funnel's spout (Deręgowski 2014). Moreover, if an insect, say a spider, were to walk from the rim of the funnel directly towards the spout, the direction of its journey would also be inverted, so that when walking downwards towards the spout it would appear to be walking upwards, but – and here is the rub – it would appears to grow smaller as it got closer to the observer. (Analogous observations apply to the spider walking towards the rim and the observer, it would on inversion appear to be moving away and yet to grow larger.) These apparent changes of perceived size are entirely contrary to the observer's daily experiences. Such violations in illusion are not unique to the example described above, with similarly nonsensical interpretations arising when we watch others walking around within Ames' room (Ittelson 1952) – they appear to grow taller or shorter as they move within it – or when we view Ames' window (de Heer & Papathomas 2017) through which a rod has been placed at right angles to the trapezoidal window and observe that as the stimulus rotates, the window and the rod that transects it appear to counter-rotate in a way that they cannot. Although the perceptual mechanisms that underlie the above examples differ, they collectively question the notion that perceptual hypotheses are direct derivatives of experience, unless it can be accepted that such derivatives may not only be facsimiles of experience, but also experiences loosely conceived as enantiomorphs. If so, then the bounds of such "enantiomorphic" regions need to be defined.

Simulation and the predictive brain

Daniel Dohrn

Dipartimento di Filosofia, Università degli Studi di Milano, 20122 Milan, Italy.
Daniel.Dohrn@unimi.it https://www.unimi.it/it/ugov/person/daniel-dohrn

doi:10.1017/S0140525X19003054, e129

Abstract

Prediction draws on both simulation and theory. I ask how simulation is defined, and what the roles of simulation and theory are, respectively. Simulation is flexible in structure and resources. Often simulation and theory are combined in prediction. The function of simulation consists of representing a situation that is relevantly like the target situation with regards to the feature predicted.

Gilead et al. transfer the opposition of theory versus simulation from the mindreading debate to prediction more generally. Against previous approaches, they emphasize that prediction may draw on both simulation *and* theory-based inference. I shall raise two interrelated issues to guide further research: (i) What is the role of simulation as distinguished from theory-based inference? (ii) How is simulation to be defined as opposed to theory?

Gilead et al. describe simulation as projecting *oneself* into a *specific spatiotemporal context*. Yet, there are doubts that simulation requires *self-projection*. Any simulation more or less detaches from one's present self. Complete detachment may be a boundary case. A relevant debate revolves around Berkeley's famous claim that imagining a tree requires imagining perceiving the tree (Berkeley (1710/1975), sect. 23; Noordhof 2002; Peacocke 1985, pp. 22–23; Williams 1973, p. 35). Yet, even if one agrees with Berkeley that simulating things requires to simulate them as they appear perceived from a spatiotemporal viewpoint, it does not follow that one projects oneself as occupying that viewpoint. In writing a detective story, we may imagine an unwitnessed murder without projecting ourselves into the murderer, the murderee, or adding a witness (Currie 1995, p. 170), though there is some debate on that (Gaut 1997). For these reasons, I also doubt the claim that the usefulness of simulation decreases the less similar simulated others are to oneself. First, one may simulate not only persons, but also things like the trajectory of a rolling boulder (Williamson 2007, p. 143). Second, overall similarity of the simulated situation to familiar environments indeed tends to facilitate simulation (Strohminger & Yli-Vakkuri 2018, pp. 318–19). Yet, this result should not be restricted to the relationship between the imagining self and simulated persons.

The imaginer may not only detach from herself, she may also largely detach from any *spatiotemporal context*. In imagining a particular shade of blue, one may leave the spatiotemporal location unspecified.

Gilead et al. are overly restrictive about the availability of "representative scenarios" to be used in simulation, for example in predicting the future divorce rates in Britain. The role of such scenarios in prediction may be highly indirect. Consider a lawyer who has handled many divorces over the decades but never tried to extrapolate general tendencies. Recent cases being more vivid in her memory, re-enacting them may lead her to judge that the divorce rate is on the rise. It may even be contested that theory would be a better guide here. The lawyer may have tacit experiential knowledge, for example, on the changing factors leading to divorce that is only retrievable by simulation (Mach 1897; Williamson 2007, pp. 145, 170).

Gilead et al. are overly restrictive either in their demands on *episodic memory*, purporting simulation to be of no avail if there are no relevant episodes (e.g., "how successful will I be as a professional wrestler?"). Yet, even if there are no pertinent *episodic memories*, one will tend to build representative scenarios from any resources available. A natural way of addressing the wrestler issue involves imagining oneself enmeshed in a wrestling match. Even if one has no remembrances of wrestling and never watched a match, one will assemble any information available in fleshing out the scene, for example, by an analogy to tavern brawls in Dutch paintings. In sum, the demarcations considered mark tendencies at best.

Coming to the choice between theory and simulation in prediction, Gilead et al. claim that simulation is more likely to be used when theory would require complex computations, cognitive resources are depleted, or when relevant socially acquired knowledge is lacking. Moreover, simulation tends to recruit less abstract representations. This suggests that simulation is as a rule less demanding in such respects than theory. The suggestion would need additional support. Theory-based inference can be easy and simple, whereas simulation can be very complicated, effortful, and resource-intensive in terms of socially acquired knowledge. It may recruit highly abstract representations. A simple folk-sociological hypothesis "nowadays people on a first date split the check" may allow me to easily derive how to act. In contrast, a psychologist's simulation of the dating situation may partly build on her past experiences and partly on arbitrarily complex empirical theories of dating behaviour. To illustrate how complicated simulation can be: Philosophers engaged in modal epistemology have us run simulations of whole *worlds*, partly by descriptive means (Chalmers 2002; Yablo 1993). I also mention computer-aided simulations in science.

In light of these considerations, I only venture a minimum characterization: *The general function of simulation consists of representing a situation that is relevantly like the target situation with regards to the feature predicted*. It is a matter of further research how to extend this minimum condition towards a full definition. Some tendencies may be observed: Simulation tends towards sensorimotor representations of spatiotemporally concrete "situation models" (Zwaan 1999; 2016), though it may recruit any mental resources, including theory (Williamson 2007, p. 143).

The decision whether to use simulation or theory depends on the informational and representational resources available to an individual. When choosing between two radio stations, I may apply a theory about which station plays classical music or simulate pleasurable experiences based on remembrances of listening to one station or the other, depending on which resource is more readily available. Complicated issues, as in the divorce rate and the wrestler example, will often elicit a case-specific *combination* of simulation and theory. It is natural to address the wrestler issue by both imagining oneself performing in a match and by drawing on theoretical knowledge (if available) about the planned routine of such events. The lawyer in the divorce rate example is well-advised to balance her tendency of simulating salient cases by consulting statistics.

In sum, the use of simulation and theory in prediction is flexible, depending on the cognitive resources available. Simulation and theory may not only be combined in prediction, they may build on each other as one resource among others. Simulation is more complex and more versatile than Gilead et al. predict it to be, and the same goes for its combination with theory.

Mind wandering as data augmentation: How mental travel supports abstraction

Myrthe Faber

Donders Centre for Cognitive Neuroimaging, Radboudumc, Kapittelweg 29, 6525 EN, Nijmegen, The Netherlands.
m.faber@donders.ru.nl

doi:10.1017/S0140525X1900311X, e130

Abstract

Gilead et al. state that abstraction supports mental travel, and that mental travel critically relies on abstraction. I propose an important addition to this theoretical framework, namely that mental travel might also support abstraction. Specifically, I argue that spontaneous mental travel (*mind wandering*), much like data augmentation in machine learning, provides variability in mental content and context necessary for abstraction.

Gilead et al. argue that the relationship between mental travel and abstraction is such that mental travel critically relies on abstraction, and that the function of abstraction is to support mental travel. I argue that in addition, mental travel – in particular, mind wandering – might facilitate abstraction, suggesting that the relationship between mental travel and abstraction as described by Gilead et al. might in fact be mutualistic. Abstraction is essential for making predictions, and critically relies on detecting invariance among experiences. For instance, based on my experience with rivers, I know that my feet get wet if I would step in one. This invariance is introduced by generalization across multiple instantiations of episodic experience, and is learnt from similarity and dissimilarity across them (Sloutsky 2003). I argue that mind wandering, which involves the spontaneous retrieval of episodic experiences, might help identify these similarities and dissimilarities, much like data augmentation in machine learning.

Mind wandering is a multidimensional construct that includes (but is not limited to) spontaneous mental travel (Christoff et al. 2016). Despite its ubiquity – rates of up to 50% of the time have been reported (Kane et al. 2007; Killingsworth & Gilbert 2010) – little is known about the function of this seemingly costly process.

In recent work, Mills, Herrera-Bennett, Christoff and I proposed that the function of mind wandering might be to support episodic efficiency and semantic abstraction (Mills et al. 2018). Specifically, we proposed the *default variability hypothesis*: mind wandering provides variability in mental content that helps to optimize the distinctiveness of episodic instantiations, which supports the extraction of invariant features of representations that ultimately lead to abstraction.

In brief, mind wandering can be characterized by its varying, dynamic content: the mind figuratively "wanders" from one thought to the next (Mills et al. 2017; 2018). These thoughts are often largely disjointed, although they might share one or more overlapping features (Faber & D'Mello 2018). Take, for instance, the following example: When reading a text about chemical properties of water, a person might think about a beach near their house, followed by a thought about their job as a beach tagger during high school, followed by a thought about a person they used to like (from Faber & D'Mello 2018). During this process, which we experience as a "train of thought," one thought likely serves as a partial cue for the next (Faber & D'Mello 2018). The process of retrieving a full memory from a partial or degraded cue is known as pattern completion, and is thought to be one of the key features of the human hippocampus (Marr 1971). Indeed, recent work has shown that the hippocampus plays a critical role in spontaneous mental time travel (McCormick et al. 2018), as its role in spontaneous retrieval of memories from partial cues facilitates mental activity that transcends the here-and-now (Faber & Mills 2018).

Importantly, the variability of content during mind wandering might support abstraction: by spontaneously retrieving a memory in a new context – either in reference to the external world or internal world – similarities across instantiations help identify the regularities necessary for abstraction. This process bears similarities with data augmentation in machine learning: diversity in data is increased without collecting new data by slightly modifying existing data, which are reused to train a model. Images, for instance, can be flipped, cropped, or partly occluded, which effectively adds noise that is useful for learning regularities across instances. A system that is learning to identify for example cats can benefit from being exposed to images that are flipped (cats can be viewed from different angles), cropped (the environment has little predictive value), or occluded (a particular feature of a specific cat might not generalize to all cats) to end up with a stable representation of "cats."

In analogy, the (re)activation of (novel combinations of) episodic experiences in the context of an unrelated or tangentially related thought or physical environment adds noise that might be useful for identifying regularities across instantiations. The data augmentation induced by mind wandering might involve noise consisting of partial activations of an experience (similar to occlusion), and/or disrupted spatiotemporal contiguity induced by the novel internal or external environment (similar to cropping). This theory suggests a potential role for mind wandering – and mental travel more generally – in facilitating abstraction through data augmentation. Adding to Gilead et al.'s theory, the ideas laid out here imply that the relationship between mental travel and abstraction is in fact mutualistic: abstraction facilitates mental travel, and mental travel facilitates abstraction. Taking into consideration the potentially mutualistic nature of this relationship is critical to understanding both mental travel and abstraction, as well as to understanding the function of the seemingly costly cognitive process of mind wandering.

The productive mind: Creativity as a source of abstract mental representations

Mark Fedyk[a] and Fei Xu[b]

[a]Betty Irene Moore School of Nursing & School of Medicine (Bioethics), University of California, Davis, Betty Irene Moore Hall, Sacramento, CA 95817 and [b]Department of Psychology, University of California, Berkeley, 3rd Floor, Berkeley Way West Building, Berkeley, CA 94720.
mfedyk@ucdavis.edu markfedyk.net
fei_xu@berkeley.edu https://psychology.berkeley.edu/people/fei-xu

doi:10.1017/S0140525X19003029, e131

Abstract

Explanations of how the brain makes successful predictions should refer to abstracta. But, the mind/brain system is for more than prediction alone. Creativity also plays an important role in supply the mind/brain system with abstracta that serve a number of valuable ends over and above prediction.

According to Gilead et al., abstracta are defined by "criteria of substitutability." They say that information out of which the mind forms dimensions along which two or more things can be substituted with one another comes from one of three sources: either the information is innate, or it is acquired from personal, subjective experience, or it is acquired from language learning and associated forms of interpersonal communication.

We believe there is a fourth source of the relevant information: the mind's creativity faculties. Some of the mind's abstracta are created – or, if you prefer, *constructed* – by the mind of the learner, rather than being derived computationally from some prior informational structure. Yet, nearly all of the by-products of the mind's creative faculties are abstracta. There is a deep connection between abstraction and creativity, therefore.

Yet, this connection is easy to overlook. Gilead et al. explain how abstraction allows the mind to leave the "here and now." The mind returns to the world by making predictions, which can then be falsified by future experience, ensuring that abstracta typically represent reality. However, this line of thinking can make it seem as if the primary function of abstracta is facilitating prediction. That is obviously an important function of abstracta – but, it is the metaphysical fact that biological organisms only move forward in time, and not a property essential to abstracta as such, which makes the connection between abstract mental representations and prediction so important.

Abstracta are for more than prediction. The brain/mind is productive, generative as often as it is predictive (cf. Fedyk & Xu 2019; Rogoff 1990; Xu 2020; Xu & Kushnir 2012), and there is probably no better example of the brain's productive capacities than creativity.

But, if creativity isn't for prediction, what is it for? We contend that two of creativity's most important functions are the facilitation of learning and the expression of acquired knowledge by making original constructs. In both cases the construction of novel abstracta is essential to creativity's ability to achieve these outcomes; some common sense examples can help clarify this claim:

- Asking questions which are not linked by any underlying logic but which generate new inquiry.
- Creating and persisting with a complex counterfactual train of thought.
- Constructing a reason why a historically trusted teacher is mistaken about a new piece of information.
- Constructing hypotheses about what ideas have not yet been considered – and doing so without carrying out an exhaustive, deterministic search of the available hypothesis space.
- Performing of a complex musical masterpiece that is original, not rote, in its performance.
- Condensing a multitude of scientific insights into a single coherent body of writing.
- Crafting a poem which almost perfectly balances form with content.
- Seeing how complex network of equations can possibly be replaced by a single equation.

In all cases, the abstractum-cum-original-construct is used for quite different purposes than prediction – and for many of these examples, a side-effect of the created abstract constructs will be increased, not decreased, surprise.

By linking abstracta with prediction, Gilead et al. are able to explain some of the normativity inherent in abstracta-based cognition: an abstractum is worth preserving in the mind's mental inventory – that is: an abstractum has epistemic value – if something in the world satisfies it, and it will therefore generally support predictions that are based upon it. But, because creativity is not for prediction, we need a different explanation of how abstracta produced by creative mental processes can have value. Our explanation of this is simple. Because the mind/brain is for more than prediction, creativity's byproducts have value when they causally facilitate any of these additional forms of value. The simplest case is when creativity facilitates the acquisition of new knowledge – for example, by inspiring unlikely explorations, questions, or curiosities. But, creativity is almost surely at the root of the construction of mental representations leading to thoughts and actions that have esthetic, mathematical, or even just hedonic value.

We, however, are particularly interested in the connection that creativity has with learning. We believe it is important to highlight the powerful compounding effect that can occur when learners are able to use creativity to deploy past learning in service of future learning. Elsewhere we have called cases where this occurs "cognitive agency" (Fedyk & Xu 2018; Fedyk et al. 2019). Relating this back to Gilead et al.'s framework, cognitive agency can be thought of as a complement to the bottom-up processes that they describe as generating abstract mental representations – cognitive agency is a top-down (or, better: top-to-top) process by which new abstracta are formed, where the new abstracta have a higher prior probability than would otherwise be the case of generating new knowledge. The concept of cognitive agency also allows us to capture the idea that it is possible for people to have a degree of control (executive function) over their learning, such that some of their decisions about learning flow partly from knowing how to learn: someone can therefore learn to learn (Lombrozo 2019), and once they know how to learn, they are potentially much more flexible in directing their efforts toward the acquisition of knowledge. And again, abstracta produced by creativity are essential for achieving this specific outcome.

Therefore, when cognitive scientists confront the question of how abstract mental entities emerge, we hope that they will include "by processes of creative thought" as among the answers. The brain subserves many different cognitive purposes: it is predictive as well as productive; creative as well as logical; symbolic as well as perceptual; and so on. Pluralism about the cognitive functions of the brain is made attractive by placing many of the considerations adduced by Gilead et al. alongside our observations about creativity. But, pluralism about the functions of the cognitive system is also an example of the flexibility that you would otherwise predict an organ like the brain to have if you knew that its capacities emerged under the forces of natural selection (West-Eberhard 1989; 2003).

Cognitive representations and the predictive brain depend heavily on the environment

Klaus Fiedler

Psychology Department, Heidelberg University, 69117 Heidelberg, Germany.
klaus.fiedler@psychologie.uni-heidelberg.de
https://www.psychologie.uni-heidelberg.de/ae/crisp/staff/fiedler.html

doi:10.1017/S0140525X1900298X, e132

Abstract

In their scholarly target article, Gilead et al. explain how abstract mental representations and the predictive brain enable prospection and time-traveling. However, their exclusive focus on intrapsychic capacities misses an important point, namely, the degree to which mind and brain are tuned by the environment. This neglected aspect of adaptive cognition is discussed and illustrated from a cognitive-ecological perspective.

Gilead et al.'s target article is no doubt written in a scholarly and eloquent way, offering conceptual common ground, and a challenge, to BBS readers from cognitive and behavioral science, philosophy, and neuroscience. Quite in the spirit of the key role attributed to abstractness in prospection and time traveling, they involve the reader in a mental travel through a conceptual landscape that is replete with abstract conceptions. The representational substrates of the mind are portrayed as a hierarchy, with manifold qualitative differences "between elements along this hierarchy, generating meaningful, often unacknowledged, diversity." In outlining this refined framework, the authors connect various approaches to cognitive psychology at a high level of construal, leaving many abstract concepts "unsaturated," open for concrete references to methods, empirical findings, and assumptions debatable at lower construal levels. The resulting portrayal is rich and unconstrained, reflecting the virtues of abstract representations with many degrees of freedom and little testable constraints on the predictive brain.

The brevity of this comment does not allow me to discuss so many distinct assumptions critically: Are modality-specific representations really based on perceptual similarity, whereas multimodal representations rely on spatiotemporal contiguity? Do categorical representations really emerge from social interaction? Rather than trying to tie down such abstract ideas to testable assumptions, I will confine myself to a critical comment on abstractness and the predictive brain as chief determinants of the adaptive ability to travel in time and space and to predict the future. This exclusive focus on intrapsychic capacities, I suspect, misses an important point, namely, the crucial role of environmental constraints imposed on cognition and the notion that the predictive environment is antecedent to the predictive brain and the predictive power of abstract mental representations. Although I anticipate that the authors may not contest but rather consider my point self-evident and common sense, the target article excludes the environment in much the same way as it was neglected during the last century of cognitive science.

Consider the authors' own prominent research program on construal-level theory (Trope & Liberman 2010), which offers a compelling explanation of abstractness (high construal level) as a key to transcending time and space (Liberman & Trope 2014). The basic idea is that different levels of abstractness are required at different levels of psychological distance. Just as a local city map represents streets and buildings in more detail at a higher resolution level than a world atlas, a letter to a close friend or partner provides more concrete references to private details than a broad publication. This natural relationship between distance (construal level) and abstractness (vs. resolution level) does not causally reflect any a-priori propensities of the mind or the brain. Instead, it is imposed by the environment. Zooming-in (reducing distance) must increase resolution and concreteness whereas zooming-out (increase distance) must reduce resolution and increase abstractness; field glasses with reverse properties would be fully dysfunctional. Likewise, the convergence of spatial, temporal, social, and evidential distance on a common distance dimension reflects constraints of the physical and social environment (Fiedler et al. 2015). What happened many years ago is more likely to have taken place at distant locations with other social partners, embedded in less likely scenarios than what is currently experienced – in the here and now.

Analogous to the primacy of physics in psychophysics, psychological distance is intrinsically entrenched in the physical and social environment, which enforces increasing abstractness (unsaturated, free parameters) when traveling to future, remote, socially unusual, or uncertain destinations. Likewise, the Weber–Fechner law (Dehaene 2003) – increasing discrimination thresholds with increasing absolute quantities – is not reversible; sensorimotor regulation would break down if thresholds for discriminating grams or milliseconds would be cruder than for discriminating tons and years.

Asymmetric mental construal mirrors asymmetries in the environment. Polarity-correspondence phenomena (Proctor & Cho 2006) reflect the alignment of memory codes with structural properties in the stimulus world, such that the presence or absence of representational features is aligned with the presence or absence of events. Research inspired by Parducci's (1965) range-frequency model shows that high-frequency categories are split into two or more subcategories, whereas low-frequency categories are merged into super-categories. Likewise, the lower density of outgroup relative to ingroup observations can explain their impoverished representation (outgroup homogeneity; Konovalova & Le Mens 2020; Linville et al. 1996). The environment confounds scarceness with distance, distance with value (Pleskac & Hertwig 2014), and novelty with uncertainty.

Not only does the environment constrain cognitive representations, it also has its own semiotic properties that determine mental representations. Sign systems can have a profound influence on information transmission and social cognition (Fiedler et al. 2008). As abstract distal concepts like risk, honesty, familiarity, or ability are not amenable to direct perception, they must be construed from vectors of proximal cues. For instance, honesty or veracity must be construed from distributive representations across such cues as pupil dilatation, gaze, disfluencies, and amount of detail. Crucially, these cues to honesty overlap considerably with the cues used to construe other distal entities, like self-confidence or social intelligence, creating semiotic confounds in mental representations. Cues to femininity overlap with cues indicating emotionality (Fiedler et al. 2008). Two personality tests, assessing extraversion and leadership, may overlap in items contents, mimicking a non-existing relation between extraversion and leadership. Cue systems offered by the semiotic environment can lead to strongly confounded cognitive representations. At different levels of abstractness, socially or culturally transmitted information takes on the properties of the sign systems used for communication.

Scientists' reluctance to reserve a place for the environment in comprehensive theories of cognition is reminiscent of the fundamental-attribution error – the preference for dispositional over situational explanations of behavior. The "predictive brain" seems to potentiate this bias, reducing the adaptive beauty of mind travel to the organism's cellular equipment. Proponents of cognitive-ecological approaches (Fiedler 2014; Pleskac & Hertwig 2014) have long complained about the neglect of extra-psychic factors, pointing out how analyzing the structure of the environment can enrich comprehensive theories of cognition, especially when it comes to prospection, time-travel, and abstraction.

Scale-free architectures support representational diversity

Chris Fields[a] and James F. Glazebrook[b]

[a]Independent, 11106 Caunes-Minervois, France; [b]Department of Mathematics and Computer Science, Eastern Illinois University, Charleston, IL 61920
fieldsres@gmail.com https://chrisfieldsresearch.com
jfglazebrook@eiu.edu https://faculty.math.illinois.edu/~glazebro/

doi:10.1017/S0140525X19002966, e133

Abstract

Gilead et al. propose an ontology of abstract representations based on folk-psychological conceptions of cognitive architecture. There is, however, no evidence that the experience of cognition reveals the architecture of cognition. Scale-free architectural models propose that cognition has the same computational architecture from sub-cellular to whole-organism scales. This scale-free architecture supports representations with diverse functions and levels of abstraction.

Gilead et al. propose an "ontology" of representation types, argue that this ontology captures "meaningful diversity in the representational substrates of the mind," and criticize the architectural assumptions of predictive-coding models as "overly simplistic" (sect. 6 paras. 2 and 4, respectively). It is never entirely clear what the elements of this ontology are – a table would have helped – but the following all seem to be included: beliefs, desires, intentions, conditions of satisfaction, subjectively distinguishable objects, features and relations represented as modality-specific, multimodal, or categorical abstractions, episodes, "lemmas" defining words, semantic and temporal networks, hierarchies, "predicators" functioning as "mentalese" verbs, models, scripts, and simulations. The distinctions between these various entities are localized to Marr's computational level of analysis (sect. 1 para. 3); however, the critique of "sub-symbolic" architectures and focus on a "layer of language like mental representations" (sect. 5.2.3, para. 7) suggest an implementation-level analysis. This distinction is critical, as few would argue that different types of representations at different levels of abstraction do not have different roles in cognition. Neuroimaging results demonstrating functional localization, for example, support functional but not architectural distinctions between types of representations.

An unstated assumption of this ontology appears to be that the structure of conscious experience is a reliable guide to the architecture of the neurocognitive system that implements this experience, including the structure of its representations. The "rich and intricate theoretical conceptualizations" that predictive-coding models are claimed to have ignored (Introduction, para. 6) are conceptualizations of the structure of a particular kind of experience, the experience of *thinking*. Hoffman (2018) and Hoffman et al. (2015) have argued on evolutionary grounds that perceptual experience is an "interface" onto the external world that supports the prediction of fitness consequences of actions but provides no reliable guide to the structure or dynamics – the architecture – of the external world. This argument can easily be inverted: conceptual experience is an interface that provides no reliable guide to the architecture of cognition. Just as humans have, in general, no need to know how computers work to use them effectively, humans have no need to know how their minds work to operate effectively in the world. A simplified folk "theory of mind" on the interface is good enough. Similar points have been made before, for example, by Chater (2018).

Relinquishing the assumption that the experience of cognition constrains the architecture of cognition is, we argue, the key to making significant progress in cognitive science. It enables asking: how is the experience of cognition produced as an *output*, and what inputs and inferential processes are needed to produce that output? Assuming the Church–Turing thesis, all computation is platform-independent: any collection of diverse representations can be generated, in principle, by any Turing-complete virtual machine. The central claim of artificial intelligence, often rendered just as "cognition is computation," is actually that cognition is *platform-independent*. It remains far from obvious, however, how to implement cognition on *any* platform, including the human brain–body system. Nor is it obvious that understanding one implementation of cognition would provide useful hints toward understanding other implementations.

The claim that cognition is scale-free is far stronger than platform independence: it is the claim that a single computational architecture works "all the way down" – describing every virtual machine at every useful level of analysis. Gilead et al. recognize this when they describe the theory of active inference, the dominant current scale-free proposal, as claiming that "the complexity of cognition can naturally arise from a canonical computation repeated across different layers of a single continuum of representational abstractness" (sect. 5.2.3., para. 3). The theory of active inference is scale-free because the free-energy principle on which it is based is scale free (Friston 2013; Friston et al. 2015), with its underlying basis, the existence of Markov blankets, derivable from classical (Kuchling et al. 2019) and even quantum (Fields and Marcianò 2019) physics. We have proposed an alternative, category-theoretic, scale-free formulation in which inferential coherence is enforced by commutativity between within-scale and between-scale mappings (Fields & Glazebrook 2019); our proposal is in the spirit of Goguen's (1991) dictum, within computer science, that abstraction *always* corresponds to the construction of a category-theoretic "cocone" as a maximal representation of inferential coherence.

Scale-free models have the advantage of being rigorously testable at every experimentally-accessible level of analysis, from those of basic physics, intracellular, and cellular processes up to the whole-organism scale and beyond. At every level, they must specify explicitly what inputs are required and what outputs are produced; indeed the role of the Markov blanket is to provide an explicit encoding of these inputs and outputs. This requirement for theoretical explicitness illuminates a key question that Gilead et al. appear to have missed: What is it about experiences of "mental travel," whether in time or across social relations, that identify them as such? What experientially distinguishes a memory from an imagined future? What distinguishes another's imagined thought from one's own? What *makes the distinction* between Gilead et al.'s "ontologically" distinct representations?

In scale-free models, such distinctions can only be made by scale-dependent *inputs*, the sources of the experienced "epistemic feelings" of reality, memory, and imagination that distinguish the functions of representations that may have the same "propositional" content. Hence, identifying these inputs is a crucially important theoretical and experimental task. Considerable progress has been made in understanding how inputs from the body and the external world are combined to locate the experienced self in the here and now (Craig 2010) and describing these processes within a predictive-coding framework (Seth & Tsakiris 2018). The signals identifying memories, future projections, and thoughts of others are less well-characterized, though it is clear that specific activities in rostral prefrontal cortex (Simons et al. 2017) and the insula – cingulate salience network (Uddin 2015) are involved. The disruption of these signals in pathology and their potential for therapeutic modulation, for example, with entheogens (Thomas et al. 2017), give their mechanistic understanding clinical urgency. Such understanding cannot be accomplished if the distinctions they signal are simply taken as given.

Representation and agency

Karl Friston

The Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, London WC1N 3AR, UK.
k.friston@ucl.ac.uk https://www.fil.ion.ucl.ac.uk/~karl/

doi:10.1017/S0140525X19002929, e134

Abstract

Gilead et al. raise some fascinating issues about representational substrates and structures in the predictive brain. This commentary drills down on a core theme in their arguments; namely, the structure of models that generate predictions. In particular, it highlights their factorial nature – both in terms of deep hierarchies over levels of abstraction and, crucially, time – and how this underwrites agency.

There are a myriad of enticing issues raised by Gilead et al. I will focus on a theme that emerges in different guises throughout their treatment. This theme is the structure of implicit generative models that the brain uses to furnish predictions of its sensorium. The nature of generative models is especially important from the perspective of active inference – a corollary of the free energy principle (Friston 2013); where many interesting aspects of generative models boil down to their *factorial* structure.

In what follows, I try to explain why generative models are so central to representation in active (Bayesian) inference as planning (Attias 2003; Baker & Tenenbaum 2014; Friston et al. 2011). I then consider the factorial nature of these models, which endows them with deep (hierarchical) structure; from the *concrete* to the *abstract* – and, crucially, from the *past* to the *future* (Friston et al. 2017d; Russek et al. 2017). Underwriting this treatment is an enactive aspect of representational processing; namely, the notion that inference about the causes of our sensations is the easy problem: the hard part is inferring the best way to gather those sensations (Davison & Murray 2002; Ferro et al. 2010; MacKay 1992).

Gilead et al. refer often to the formalism of active inference. I think this is perfectly appropriate, because a formal treatment of representational structure is, in its essence, a treatment of the generative models that underwrite inference. Technically, a generative model is just a probability distribution over some causes and their consequences. In the setting of the embodied brain, the causes are states of the world "out there" – that are *hidden* behind our sensations. These sensations are the consequences. Inverting a generative model refers to the inverse mapping from (sensory) consequences to their (worldly) causes. These causes are the *abstracta* and *concreta* that constitute different kinds of representations in Gilead et al. The generative model is important because most of the heavy lifting – in terms of understanding structure–function relationships in the brain – rests on its form. In other words, if one knows the generative model, model inversion can be cast in terms of the Bayesian brain hypothesis (in a normative sense) (Doya 2007; Knill & Pouget 2004) or combined with standard inversion schemes to generate neuronal processes theories about computational brain architectures and neuronal message passing (Friston et al. 2017c).

These theories are usually cast in terms of belief-updating *via* a gradient descent on variational free energy. There are several schemes that fall under this class; all of which have been used as biologically plausible process theories for perceptual inference. Crucially, exactly the same quantity is optimised by action; thereby providing a formal account of the action–perception cycle (Fuster 2004). Particular instances include predictive coding (Rao & Ballard 1999) and variational message passing for generative models based upon continuous and discrete states, respectively. These process theories constitute a field in cognitive neuroscience that has become known as predictive processing (Clark 2013; Seth 2014). Therefore, what are the most important aspects of a generative model?

One aspect has already been mentioned; namely, the distinction between continuous and discrete models. However, a feature that is common to both is their factorial structure. In fact, from a technical perspective, the way in which we factorise our (non-propositional) posterior beliefs about hidden causes (i.e., how we come to represent things "out there") rests upon a factorisation known as a *mean field approximation* in physics and machine learning. Key examples emerge throughout (Gilead et al.). The first is a factorisation over the levels of a deep (hierarchical) generative model. Typically, the lowest levels – that generate sensory data – are concrete and modality bound. As one ascends the hierarchy, the states of the world represented become more abstract and inclusive.

Another important aspect of factorisation is a carving of putative hidden states of the world within any hierarchical level. My favourite example is the factorisation into "what" and "where" (Ungerleider & Haxby 1994). In short, knowing what something is does not tell you where it is and *vice versa*. This (conditional) independence is manifest beautifully, in terms of the functional anatomy of the dorsal and ventral streams in the brain. This sort of factorisation emerges frequently in Gilead et al. One intriguing example is the notion of predicators; namely, *representations that behave like functions*. An interesting question here is whether one needs to treat relationships in a way that is fundamentally different from objects? For example, how is a representation of "what" formerly distinct from a representation of "where," when generating visual input?

The final factorisation is over time. This theme emerges in modality-specific features, objects, and relationships – that rest upon the notion of *object permanence*. This sort of permanence has to be written into a generative model of a capricious world. This theme reappears in terms of *spatiotemporal contiguity* in the treatment of multimodal features. Indeed, the premise of Gilead et al. rests upon integrating influential theories in the "predictive brain camp" with "prospection (or future oriented mental time travel)." This is a big move, because it entails generative models of dynamics, narratives, or trajectories – with representations of the past and future. In turn, this enables the representation of states that have not yet been realised. These states undergird "simulation of future events" (intro., para. 4) and a sense of agency. In other words, the notion of a model that can "generate a representation that models the specific problem at hand" (sect. 3.1, para. 3) is exactly a generative model of the future, with "my" action as a latent state that has to be inferred. This is important from the point of view of active inference, because it suggests that much of our inference is not about states of affairs "out there" but more about "what would happen if I did that" (Schmidhuber 2006). This is nicely summarised as (ibid., p. 23):

> The functionality of a simulation stems from the fact that the person running the simulation self-projects into it, that is, becomes an agent in the simulated situation.

Gilead et al. then offer a compelling conclusion about mental travel (sect. 4, para. 2):

> Representational structures … form the bridges that allow us to traverse uncertainty. In light of this … the link between abstraction and mental

https://doi.org/10.1017/S0140525X19002000

travel is fundamental to any consideration of these constructs; there is no mental travel without abstraction, and there is no need for abstraction but to support mental travel.

I would add:

and there is no need for mental travel that but to support inference about what I should do next.

Structured event complexes are the primary representation in the human prefrontal cortex

Jordan Grafman[a,b,c]

[a]Cognitive Neuroscience Laboratory, Shirley Ryan AbilityLab, Think and Speak Lab, Chicago, IL 60011-5146; [b]Departments of Physical Medicine and Rehabilitation, Neurology, Psychiatry, and Alzheimer's Center and Cognitive Neurology, Northwestern University School of Medicine, Chicago, IL 60611 and [c]Department of Psychology, Weinberg School of Arts and Sciences, Evanston, IL 60208.
jgrafman@northwestern.edu

doi:10.1017/S0140525X19003200, e135

Abstract

Instead of endorsing an all-encompassing view about the influence of abstractions in predictive processing, I suggest that most deliberative thought including complex abstractions, agent actions, and/or perceived environmental sequences are stored in the human prefrontal cortex in the form of structured event complexes.

Gilead and colleagues propose that abstraction underlies the predictive nature of cognition. Functional neuroimaging and lesion mapping studies indicate that abstraction and long-term prediction are both associated with the human prefrontal cortex (HPFC). I suggest that the HPFC is also a good fit for the storage and use of *time- and event-related higher order representational knowledge* – contributing to predictive processing beyond a single event (Wood & Grafman 2003). I have no dispute with Gilead et al. using the term *hierarchical* to describe the dominant role of the HPFC in storing, organizing, controlling, and issuing the intentional commands to execute most daily routines or adaptive behavior. Although this is by no means a new idea (e.g., see Miller et al. 1986; Schank & Abelson 1977), the recent development of statistical and computational modeling and information theory tools Gilead et al. describe have substantially improved the quantification of hierarchical models and their predictions.

The environment that surrounds us contains events of various durations. One possible way humans have come to represent events of varying durations is by parsing the brain into differing representational sectors with each sector containing representations of event series of differing durations (Radvansky & Zacks 2017). This is a hierarchical schema, with increasing number of events composed of longer durations more likely to be stored in the anterior sectors of the HPFC. Prediction would require retrieving such stored events held in long-term memory (Grafman et al. 2005). Exemplar features such as complexity, abstractness, frequency of exposure, and relational similarity could emerge from the organizational structure of these representational networks. Flexible parsing rules (see Zacks et al. 2016) allowed events to be represented in structured event complexes. This organizational principle could facilitate switching between complex representations and abbreviated heuristic knowledge contained in posterior cortical/subcortical networks that store more time-compact single event representations conserving the brain's free energy (Parr & Friston 2019).

These HPFC structured event complexes including plans, narratives, abstractions, deliberations, and goal-derived activities would be stored in networks along the same principles that enable words or objects to be stored in a coherent cognitive network. This would, for example, allow for prioritizing structured event complexes performed frequently in the real world by giving their representations cortical space distinctiveness and activation default superiority (Grafman et al. 1991; Rosen et al. 2003). Although this would be true of any stored memory (e.g., words or objects), it is particularly crucial for sustaining the activation of longer duration memory representations that support goal achievement and social navigation. But, why couldn't our brain just be organized around short time duration representations that were dynamically and adaptively repackaged into a sequence depending upon intended or environmentally provoked behavior? That would require neural resources that would be continually engaged in such activity causing laborious multi-tasking, siphoning the brain's energy deposits and cognitive resources.

Using functional neuroimaging, Etienne Koechlin and David Badre have demonstrated that distinct brain regions may support different levels of a hierarchical processing framework (Badre & Nee 2018; Koechlin & Jubault 2006). Although these authors have focused mostly on left dorsolateral prefrontal cortex language mechanisms, their studies have demonstrated the feasibility of using a mathematically constrained hierarchical model to predict differing levels of processing and representation. Other functional neuroimaging and lesion mapping studies of chess players (Nichelli et al. 1994), processing narratives (Nichelli et al. 1995), and script decision-making (Sirigu et al. 1996) have supported the structured event complex (SEC) conceptualization.

But, how are durations first encoded and later parsed? There is some evidence that individual segments or events are first tabulated in childhood, later compiled into sequences that can then be reduced to heuristics and stored posteriorly or deeper in the brain (Rattermann et al. 2001). This hierarchy of duration representation doesn't mean one form replaces another as we mature and learn. Rather, it is likely that all of these forms of representations remain stored across the HPFC with various representational forms activated depending on situational needs.

Cautions

A challenge for comparing hierarchical forms of knowledge to see if they occupy distinct brain sectors is to make sure the compared tasks are psychometrically matched. For example, controlling for the difficulty in retrieving different categories of words based on frequency or age of acquisition of the word is a magnitude easier than controlling for the difficulty level of abstract tasks or event narratives. Chapman and Chapman described the problems and solutions in creating tasks that could be psychometrically comparable when focusing on dissociations decades ago (Chapman & Chapman 1978; 2001). Also note that much of the literature Gilead et al. cite depends on correlational functional neuroimaging findings in healthy volunteers rather than using causal lesion mapping or non-invasive brain stimulation results. Providing convergent data from studies using different techniques to muster support for a specific idea or theory is important.

To summarize, a complementary perspective to Gilead and colleagues' all-encompassing view about the influence of abstractions in predictive processing is that the HPFC evolved to capture events that occur over longer and longer periods of time. All deliberative thought including complex abstractions, social agency, and narrative explanations that fit within structured event complexes devoted to representing information occurring over multiple events and time periods will be represented in the HPFC. Such HPFC representations will bind via a variety of neural mechanisms to representations stored elsewhere in the brain providing a relatively complete *network-based* capture of all the information composing that time frame. The large number of exemplars within each hierarchical SEC network would allow abstract representations to emerge that capture the similarities *across* representational exemplars, but preserve the individuated representations that are critical to remembering episodes. Given that longer time frames impose greater persistence toward the completion of an activity and achievement of a goal, activation of one or more structured event complexes inhibit diversions enabling reinforced and rewarded goal-directed activity in the face of environmental distractors. Other species cannot compete with humans because of the richness and multiplicity of our structured event complexes. Forming predictive abstractions are dependent on the existence of structured event complexes.

Experiences of liking versus ideas about liking

Alison Ledgerwood[a], Paul W. Eastwick[a] and Bertram Gawronski[b]

[a]Department of Psychology, University of California, Davis, CA 95616 and
[b]Department of Psychology, University of Texas at Austin, Austin, TX 78712
aledgerwood@ucdavis.edu http://www.alisonledgerwood.com/
eastwick@ucdavis.edu http://pauleastwick.com/
gawronski@utexas.edu http://www.bertramgawronski.com/index.htm

doi:10.1017/S0140525X19003145, e136

Abstract

We leverage the notion that abstraction enables prediction to generate novel insights and hypotheses for the literatures on attitudes and mate preferences. We suggest that ideas about liking (e.g., evaluations of categories or overall traits) are more abstract than experiences of liking (e.g., evaluations of particular exemplars), and that ideas about liking may facilitate mental travel beyond the here-and-now.

Gilead et al. propose that abstract representation enables predictive cognition. Applying this idea to two major areas in social psychology – attitudes and close relationships – generates novel insights and hypotheses for the science of human evaluation and interpersonal liking. In particular, Gilead et al.'s framework points to an important distinction between *experienced* evaluations (e.g., "Right now, I like this tall man") and abstract *ideas* about liking (e.g., "Generally, I like tall men"). Furthermore, their arguments suggest that abstract ideas about liking enable predictive cognition and mental travel.

Attitude researchers and close relationships researchers have not directly made the distinction between experienced evaluations versus abstract ideas about liking. In this commentary, we leverage Gilead et al.'s framework to highlight important new directions for each research area (Fig. 1).

First, attitude researchers often study liking for social and nonsocial categories using measures that focus on liking for the overall category (e.g., a person's evaluation of the category "African Americans") as well as measures that focus on liking for individual exemplars of that category (e.g., the average of a person's evaluations of a series of individual African American faces). In Gilead et al.'s language, evaluations of categories versus exemplars can be arranged along a continuum of mental abstraction: People's ideas about how much they like broad social categories are more abstract than their experienced evaluations of specific exemplars. Yet researchers have typically treated evaluations of exemplars and evaluations of categories as conceptually equivalent (see e.g., Ajzen & Fishbein 1977, who treated both as measures of so-called "general attitudes;" see Cooley & Payne 2019, for a notable exception), which seems problematic in light of Gilead et al.'s framework. For example, Gawronski (2019) pointed out that the literature on implicit bias has largely ignored confounds between type of measure (implicit versus explicit) and target object (exemplars versus categories). Whereas explicit bias measures typically involve evaluating categories, implicit bias measures involve evaluative responses to exemplars.

Gilead et al.'s arguments further suggest that category evaluations, like other abstract representations, may function to enable predictive cognition. Whereas experienced evaluations of specific exemplars may guide immediate decisions about what to do in the here-and-now, abstract evaluations of overall social categories may guide decisions about what to do at a spatially distant location, in a hypothetical scenario, or in the future (e.g., who to hire to fill a future position). Moreover, abstract evaluations may be especially useful for making predictions about situations one has not yet experienced and that may be difficult to simulate (e.g., whether to move to a new city with a particular set of demographics).

Second, human mating researchers often study liking for attributes using measures that focus on liking for an overall attribute (e.g., a person's evaluation of the trait "intelligence" in a romantic partner) as well as measures that focus on liking for individual exemplars that vary in terms of a given attribute (e.g., the extent to which intelligence in a series of individual potential partners drives a person's evaluation of each partner). Again, these evaluations can be arranged along a continuum of mental abstraction. People's ideas about how much they like an attribute as a general concept are more abstract than their experienced evaluations of specific exemplars that embody those attributes. Yet researchers have typically treated them as conceptually equivalent (for reviews, see Eastwick et al. 2014; Ledgerwood et al. 2018), which again seems problematic in light of Gilead et al.'s framework. For example, Ledgerwood et al. (2018) observed that the literature on human mating has largely ignored the distinction between evaluative experiences of traits (e.g., people's evaluations of romantic partners who are more vs. less intelligent) and evaluations of traits in the abstract (e.g., people's evaluations of the trait *intelligence* in a partner).

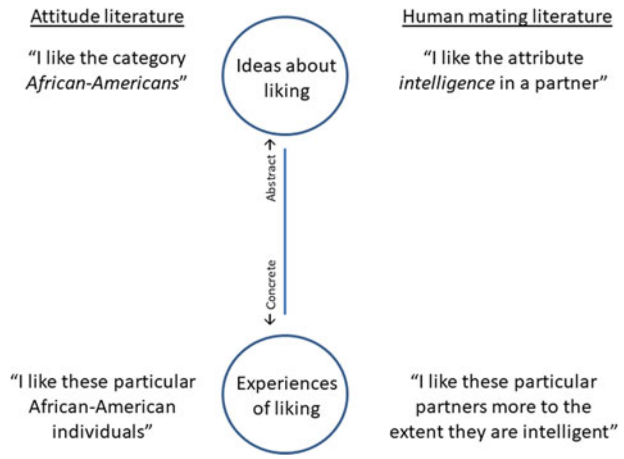

Figure 1. (Ledgerwood et al.) Liking in two literatures. Ideas about liking (top) are more abstract than experiences of liking (bottom) because they treat individual exemplars as substitutable.

Distinguishing between abstract and concrete attribute evaluations has allowed us to ask new questions about how people form abstract representations of liking for attributes. For example, our research suggests that people form abstract attribute preferences by drawing on not only their concrete evaluative experiences, but also incidental features of the learning context (Eastwick et al. 2019; Wang et al. 2020). These incidental contextual features include how plentiful a trait is (e.g., whether the potential mates that one encounters are generally high or low in *intelligence*) and how much liking someone is generally experiencing (e.g., whether the potential mates that one encounters are generally desirable or undesirable). Thus, people's ideas about how much they like various traits in the abstract may be biased by the context in which they learn about their likes and dislikes.

Furthermore, abstract attribute preferences, like abstract category evaluations, may serve the critical purpose of enabling predictive cognition. For example, humans may rely on their abstract attribute preferences to predict whether they will like a potential date (or friend or colleague) who is spatially distant rather than close, or hypothetical rather than real (see Eastwick et al. 2011; Huang et al. 2020). Moreover, abstract attribute preferences may be especially useful for making predictions about situations that one learns about through socially acquired knowledge, rather than direct experience, and that may therefore be difficult to simulate (e.g., whether to visit a bar that a friend describes as full of particularly quirky patrons). Consistent with this notion, recent research suggests that abstract, summarized preferences primarily predict situation selection at a distance (e.g., whether to sign up for a dating website described as featuring highly intelligent partners) rather than situations that have been directly experienced (e.g., whether to sign up for a dating website after experiencing example profiles of potential partners that look highly intelligent; Wang et al. 2020).

Going forward, we urge scholars to more seriously distinguish between abstract ideas about liking and concrete experiences of liking. If abstraction enables predictive cognition, as Gilead et al. posit, this distinction may prove both crucial and generative.

Prospection does not imply predictive processing

Piotr Litwin[a,b] and Marcin Miłkowski[b]

[a]Faculty of Psychology, University of Warsaw, 00-189 Warszawa, Poland and [b]Institute of Philosophy and Sociology, Polish Academy of Sciences, 00-330 Warszawa, Poland.
piotr.litwin@psych.uw.edu.pl
mmilkows@ifispan.waw.pl
http://marcinmilkowski.pl/en/

doi:10.1017/S0140525X19002991, e137

Abstract

Predictive processing models of psychopathologies are not explanatorily consistent with the present account of abstract thought. These models are based on latent variables probabilistically mapping the structure of the world. As such, they cannot be informed by representational ontology based on mental objects and states. What actually is the case is merely some terminological affinity between subjective and informational uncertainty.

Gilead et al. propose an ontology of mental representations and argue that it may have important implications for predictive processing models of psychiatric disorders. Indeed, the proposed account and predictive processing seem to share some similarities: In both, representations form a hierarchical, tree-like structure organized along the continuum of abstractness, and this organization serves the goal of uncertainty mitigation.

Unfortunately, the authors do not specify how exactly their representational account could inform predictive processing models of psychopathology – and these models could certainly use some guidance. Let us consider models of schizophrenia. Phenomena such as diminished oddball effects or susceptibility to visual illusions were proposed to arise from the failure to attenuate sensory precision (Friston et al. 2016), that is, weak low-level (perceptual) priors (Sterzer et al. 2018), which fail to constrain sensory input in accordance with the brain's expectations. Nonetheless, hallucinations should emerge because of strong priors exerting disproportionate influence on perceptual inference, producing percepts out of thin air (Corlett et al. 2019). Low-level priors were found to be both *reduced* in delusion-prone patients (Schmack et al. 2015) and *enhanced* in hallucination-prone healthy individuals (Powers et al. 2017). Even more perplexingly, recent studies showed that the derivation of perceptual priors from natural scene statistics (Kaliuzhna et al. 2019) or the susceptibility to a wider range of visual illusions (Grzeczkowski et al. 2018) is actually unaffected in schizophrenic individuals.

These glaring inconsistencies tend to be explained as disturbances in global predictive dynamics. Accordingly, inefficient lower-level priors are compensated by higher-order, semantic prior beliefs, simultaneously driving hallucinations through the facilitation of sensory activations consistent with delusional beliefs (Corlett et al. 2019; Sterzer et al. 2018; 2019). However, this contradicts the core assumption of predictive processing that all cognitive processes arise from computations performed at various levels of a *single*, *homogeneous* representational hierarchy, with only *adjacent* layers interacting directly (Williams 2018). The assumption of adjacency is shared by virtually all contemporary computational psychiatry models, regardless of their exact implementation – be it a Deep Boltzmann Machine (Corlett et al. 2019), belief propagation algorithm (Denève & Jardri 2016), or hierarchical Bayesian inference as envisioned by vanilla predictive processing. Altered global dynamics could possibly give rise to phenomena observed in psychiatry, but it does not entail unmediated interaction between non-adjacent layers, as it cannot occur in hierarchical architectures. Finally, it is rather unlikely that higher-order priors could enhance signaling in sensory cortices. Top-down connections are taken to be inhibitory rather than excitatory in predictive processing (Denève & Jardri 2016), as their main job is to suppress prediction errors. Thus, predictive processing accounts of schizophrenia themselves seem to be like the robot Herbie (mentioned in the article, p. 52), as they "cannot abide by some of [their] imperatives without breaking others."

We actually believe that the representational hierarchy introduced by Gilead et al. could alleviate some of the problems that haunt predictive processing models of psychopathology. The account specifies relations (e.g., the relative position in the hierarchy) between abstracta and the degree of "abstraction saturation" of representations. Thus, it provides a clear definition of how continuum of abstractness orders the hierarchy, which remains an unsolved problem for predictive coders (see Williams 2018). However, how could it discharge the inconsistencies discussed above? The proposed hierarchy of representations is expressed in classic ontological categories of mental objects and states (and interactions between these symbolic representations). It is unclear how it could inform a sophisticated take on representation in predictive processing: It is assumed to resemble the causal-probabilistic structure of the environment in the form of latent probabilistic variables occupying various levels of inferential hierarchy (Gładziejewski 2015). The causal matrix of external causes of a system's activations is represented probabilistically in the form of posterior distributions determining prior distributions on plausible parameter (posterior) values at the subordinate level. One cannot even discuss the exact problems of predictive processing without this probabilistic language, not to mention providing ailments for them. In particular, some of the proposed abstract representations cannot be easily placed in a single, homogeneous hierarchy (e.g., predicators which serve as representational entities to be filled by other abstract representations).

Gilead et al. advance the case of "scruffies," arguing that both kinds of representations (symbolic and structural/probabilistic) may play a significant role in cognition. We see it becoming more common in the debate, as if being a "scruffy" absolved one of the sins of providing disconnected or piecemeal explanations. Yet, the presented account remains silent about how symbolic and subsymbolic representations could interact. A few simple tricks, such as (1) a recourse to higher-order processes (conceptualization, inner speech, or communication, which were traditionally taken to arise from symbolic processes), (2) rephrasing them in predictive processing terms, and (3) providing a Bayesian "just-so" story on how emerging beliefs may lead to psychological transformation, will not do. They are not enough to show that the account can guide further development of predictive processing models in psychiatry.

Thus, we consider the "important implications" for predictive processing merely declarative and founded on mere terminological similarities between *prospection* and *prediction*. The lure of equivocation is strong, but there are actually major differences that cannot be overlooked. The presented account focuses on *particular cognitive phenomena* (e.g., future-oriented mental time travel), proposing how representational, abstractness-organized hierarchy allows *humans* to mitigate *subjective*, *emotionally-laden* uncertainty stemming from psychological distance. In contrast, predictive processing claims that the *brains* attempt to mitigate *informational* uncertainty, and that computational processes that serve this purpose underlie *all cognition*. But, why connect these accounts? In our opinion, one does not have to relate to a dominant explanatory paradigm at all costs in order to justify one's own account's explanatory potential/value. We believe that the presented theory is easily self-standing.

Acknowledgments. The authors were supported by the National Science Centre (Poland) research grant under the decision DEC-2014/14/E/HS1/00803.

Catching the intangible: a role for emotion?

Maria Montefinese[a], Ettore Ambrosini[a,b], Antonino Visalli[a] and David Vinson[c]

[a]Department of General Psychology, University of Padova, 35131 Padova, Italy; [b]Department of Neuroscience, University of Padova, 35128 Padova, Italy and [c]Department of Experimental Psychology, University College London, London WC1E 6BT, UK.
maria.montefinese@unipd.it https://sites.google.com/view/mariamontefinese
ettore.ambrosini@unipd.it
https://www.researchgate.net/profile/Ettore_Ambrosini
antonino.visalli@unipd.it
https://www.researchgate.net/profile/Antonino_Visalli
d.vinson@ucl.ac.uk
https://www.ucl.ac.uk/pals/research/experimental-psychology/person/david-vinson/

doi:10.1017/S0140525X19002978, e138

Abstract

A crucial aspect of Gilead and colleagues' ontology is the dichotomy between tangible and intangible representations, but the latter remains rather ill-defined. We propose a fundamental role for interoceptive experience and the statistical distribution of entities in language, especially for intangible representations, that we believe Gilead and colleagues' ontology needs to incorporate.

In the spirit of the predictive nature of cognition, we agree with Gilead and colleagues that a predictive brain framework for abstract representations, contemplated as a hierarchy ranging from the tangible to the intangible, could be salutary. However, it is important to recognize that although a crucial aspect of the ontology proposed by Gilead and colleagues is the dichotomy between tangible and intangible entities, the latter remains rather ill-defined despite the formal treatment (sect. 2.1, para. 3). In particular, Gilead and colleagues define "intangible abstracta" (often called "abstract representations/concepts" in the literature on semantic representations) as categories whose concreta are not detected by our senses, but mainly transmitted from mind to mind using language. However, they also propose that some intangible dimensions of the intangible abstracta "may have an *innate* basis, or may be emergent properties discovered via *personal experience*" (sect. 2.1, para. 3), properties also relevant for the modality-specific and multimodal abstracta (both based on sensorimotor features) (sect. 2.1, para. 1 and 2). Consequently, the distinctions between the different kinds of representations are obscure and Gilead and colleagues' definition of "intangible abstracta" seems somewhat contradictory to us. Therefore, it is important to get a clear idea of how personal experience and social interaction combine to produce intangible abstracta.

In light of these theoretical considerations, we propose that many intangible representations could be intangible abstracta with affective content. The plausibility of this view has been supported by many studies demonstrating the crucial role of emotion for intangible abstracta (Crutch et al. 2013; Kousta et al. 2011). In particular, although tangible entities have direct sensory referents (Crutch & Warrington 2004; Montefinese et al. 2013; Paivio 1971), intangible abstracta tend to be more emotionally valenced (Crutch et al. 2013; Kousta et al. 2011; Vigliocco et al. 2013) and have low sensorimotor grounding (for a concise review, see Montefinese 2019). In line with the idea that affective content is particularly relevant for intangible abstracta representation, a number of neuroimaging studies showed that intangible abstracta processing increases activation in brain regions involved in emotion processing (Vigliocco et al. 2013; Wang et al. 2018), such as the rostral anterior cingulate cortex.

Very recently it has been proposed that *interoception* (the perception of the internal state of the body) contributes to the perceptual grounding of intangible abstracta. Crucially, interoception is the most important perceptual modality in the experience of emotions, especially the negative ones (e.g., fear and sadness), over and above the traditional five sensory modalities (Connell et al. 2018). An exploration of emotion and of its perceptual grounding via interoception seems like a necessary step in building a comprehensive theory of abstract representational capacities.

Still, taking affective information into account might not suffice to capture representation of intangible abstracta. In this regard, recent multimodal models suggest that supplementing affective information with information related to the statistical distribution of concepts in language (i.e., *distributional models* of semantic representation; Landauer & Dumais 1997) drastically improves prediction of human affective judgments (Bestgen & Vincze 2012; Recchia & Louwerse 2015; Vankrunkelsven et al. 2018). More importantly, recent work by Lenci et al. (2018) reveals a strong link between distributional statistics and emotion: *intangible representations* have more affective content and tend to co-occur with contexts with higher emotive value. However, it is worth noting that the contribution of the distributional models to semantic representation goes beyond that of affective intangible abstracta. Indeed, it has been shown that these models can successfully account for semantic and linguistic judgments, as well as higher-level judgments such as probability judgments and risk perception in a human-like manner (Bhatia et al. 2019; Rotaru et al. 2018). As is the case for emotion, the importance of distributional information for intangible abstracta is also supported by neuroimaging studies. Intangible abstracta reliably engage neural systems associated with linguistic processing (especially, left anterior temporal cortex and left inferior frontal gyrus) to a greater extent than tangible abstracta (Wang et al. 2010). Increased activity for intangible abstracta in networks associated with language processing appears to be specifically associated with distributional similarity, versus other aspects of intangible representations which do not appear to be localized to language-related networks (Wang et al. 2018). As intangible abstracta are mainly acquired through verbal experience (as Gilead and colleagues acknowledge in sect. 2.1, para. 3) and the distributional theory represents one of the main theoretical frameworks in the semantic literature, it is surprising that such a role of language is not addressed directly. Given the importance of these models in explaining both intangible and tangible representations, we think that Gilead and colleagues should incorporate them in their theory. Moreover, by revealing the statistical relations between abstract entities, distributional models represent a powerful tool to integrate Gilead and colleagues' account and predictive brain theories, which assume the brain as a statistical inferential machine.

In short, we believe that Gilead and colleagues have missed a chance to "provide cognitive scientists with an *accurate* ontology of the representational entities that exist in our mind – and that subserve predictive cognition" (sect. 5.1, para. 5). What we think is missing from their analysis is how emotion and distributional information fits in with the proposed ontology. In keeping with a metaphor used by the authors, interoceptive experience and linguistic distribution would represent two additional "tricks" used by our brain both to build the different layers of the representational hierarchy and to "transcend the here-and-now," and we think that the authors' model could benefit from integrating them.

Above and beyond the content: Feelings influence mental simulations

Kellen Mrkva[a], Luca Cian[b] and Leaf Van Boven[c]

[a]Columbia Business School, Columbia University, New York, NY 10027; [b]Darden School of Business, University of Virginia, Charlottesville, VA, 22903 and [c]University of Colorado Boulder, Boulder, CO 80305.
km3386@columbia.edu www.kellenmrkva.wordpress.com
CianL@darden.virginia.edu
https://www.darden.virginia.edu/faculty-research/directory/luca-cian
vanboven@colorado.edu
https://sites.google.com/colorado.edu/leafvanboven

doi:10.1017/S0140525X19003108, e139

Abstract

Gilead et al. present a rich account of abstraction. Though the account describes several elements which influence mental representation, it is worth also delineating how feelings, such as fluency and emotion, influence mental simulation. Additionally, though past experience can sometimes make simulations more accurate and worthwhile (as Gilead et al. suggest), many systematic prediction errors persist despite substantial experience.

Gilead et al. describe a new, exciting theory of abstraction, which is a step forward for work on this interdisciplinary topic. Among other contributions, their theory will allow researchers across fields to use a common terminology and better appreciate insights from other fields.

Two key elements that need to be integrated into their account are (1) an appreciation of how feelings such as fluency influence mental simulation, and (2) an appreciation of how prediction errors persist even among people with substantial personal experience.

Both of these elements could be integrated into Gilead and colleagues' account, though it would be worthwhile delineating how.

Gilead et al. argue that people construe objects abstractly when these objects are psychologically distant and when it is rational to do so based on detail/accuracy trade-offs. Yet, feelings such as fluency also influence construal level and mental simulation. People are more likely to simulate events when they come to mind easily (Kahneman & Tversky 1982; Schwarz et al. 1991).

Feelings such as fluency and mood influence mental construal. In everyday life, people attend to their feelings as a source of information, and different feelings provide different types of information and trigger different processing strategies (Schwarz & Clore 1996). For example, being in a negative (vs. positive) mood can spontaneously generate a detail-oriented elaboration, which is usually adaptive in problematic situations (Schwarz 2001). A negative mood, by signaling not only a problem but also its imminence, can lead people to adopt a concrete construal and focus attention on the immediate, which can increase how much immediate events are prioritized (Labroo & Patrick 2009; Mrkva & Van Boven 2017). In contrast, positive mood, by signaling that the immediate environment is benign, encourages an abstract construal and a long-term perspective. Further findings (see, e.g., Bless et al. 1996; Gardner et al. 2014) corroborate the idea that a negative mood elicits proximal, concrete construals whereas a positive mood often elicits distal, abstract construals. In addition to positive or negative mood, both emotional arousal and fluency reduce feelings of psychological distance and influence mental simulations (Alter & Oppenheimer 2009; Mrkva et al. 2018; Szpunar & Schacter 2013; Van Boven et al. 2010). Bodily sensations, such as upward or downward head and eye movements can also influence simulations and mental construals, as when looking upward evokes more abstract imagery of distant objects (Barsalou 1999; Cian 2017; Van Kerckhove et al. 2014).

Because feelings such as fluency influence mental simulation, people's feelings can make them more likely to simulate events even when fluent events are objectively less likely to occur. For example, people might simulate an extremely unlikely terrorist attack when it is top-of-mind and fluent, whereas failing to simulate more likely risks for which they have equivalent information and experience (Mrkva et al. 2018; Sherman et al. 1985; Sunstein 2003). Doing so can have drastic consequences, as when Americans decided to drive rather than fly shortly after 9/11, likely resulting in hundreds of additional deaths from traffic accidents (Blalock et al. 2009).

Finally, feelings can lead to empathy gaps and other systematic prediction errors. Empathy gaps prevent people from effectively simulating other mental states, even when they possess abundant relevant conceptual knowledge. For example, when in neutral or satiated states, students do not appreciate how much hunger would influence preferences (Van Boven & Loewenstein 2003), smokers fail to appreciate how strong their desires to smoke will be (Sayette et al. 2008), and drug addicts underestimate how alluring their urges will be (Van Boven et al. 2013). These empathy gaps persist even amid substantial personal experience, considering that smokers and drug users typically have vast experience with the relevant concreta and abstracta (Van Boven et al. 2013). They also occur even when people in neutral "cold" states and emotional "hot" states have the same objective contextual details about an event (Van Boven et al. 2013). In particular, people may have difficulty translating experiences in an emotional, "hot" state to an abstract lesson implemented in a "cold" state (and vice versa; Loewenstein 1996; Loewenstein et al. 2003; Van Boven & Loewenstein 2003).

Another systematic prediction error, the planning fallacy, persists in spite of substantial personal experience (Buehler et al. 1994; 2002). People underestimate how long tasks will take to complete and how much money they will spend, even if they are experienced with these decisions and have made many similar underestimation errors in the past (Buehler et al. 2002; Griffin & Buehler 2005; Peetz & Buehler 2009). These errors partly reflect tendencies to rely on hope over experience, simulate the easiest-to-imagine, best-case scenario, and ignore other tasks and obstacles that may interfere (Buehler et al. 2002; Buehler & Griffin 2003; Kahneman & Lovallo 1993). Gilead et al. state that people should (and do) simulate events when they have past experience in their reservoir of memories. Yet, it is important to note that people are prone to systematic prediction errors such as empathy gaps and the planning fallacy despite having previous experience. The observation that empathy gaps and other prediction-reality gaps persist even among people with vast experience offers challenges for any predictive cognition framework that focuses exclusively on experience and conceptual knowledge.

In sum, Gilead et al. develop a rich framework, but need to allow ample room in their framework for feelings (emotions, fluency, and other states). These feelings can shape whether people use abstract thought, how they are impacted by simulations, and whether their predictions will correspond to reality.

Are all distances created equal? Insights from developmental psychology

Bronwyn O'Brien, Joshua L. Rutt
and Cristina M. Atance

Faculty of Social Sciences, School of Psychology, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5.
bobri082@uottawa.ca
jrutt@uottawa.ca
atance@uottawa.ca
http://www.socialsciences.uottawa.ca/ccll/eng/

doi:10.1017/S0140525X19003078, e140

Abstract

Gilead et al.'s theory presupposes that traversing temporal, spatial, social, and hypothetical distances are largely interchangeable acts of mental travel that co-occur in human ontogeny. Yet, this claim is at odds with recent developmental data suggesting that children's reasoning is differentially affected by the dimension which they must traverse, and that different representational abilities underlie travel across different dimensions.

We appreciate Gilead et al.'s highlighting of the various "epistemic barriers" or "dimensions" (in addition to "time") that humans must mentally traverse. Inherent to their theory of mental travel is that the different dimensions/distances (i.e., temporal, spatial, social, and hypotheticality) are intertwined, mostly interchangeable (thus lumped together under the broader construct of "mental travel"), and co-occurring in ontogeny. However, this characterization overlooks important nuances between different forms of mental travel and the diverse representational capacities subserving them. This is especially apparent in recent cognitive developmental data suggesting that children's sensitivity to different dimensions emerges in a staggered manner, and that the capacity to traverse dimensions may rely on different underlying representational capacities.

Developmental research has not yet provided a comprehensive account of when different forms of "mental travel" emerge and, importantly, whether sensitivity to each dimension emerges simultaneously. Yet, new data suggest important differences as a function of the dimension children must traverse. A notable point is that traversing social distance appears to emerge early in development. For example, taking a socially distanced perspective (by reasoning about "other" vs. "self") enhances 3- to 5-year-olds' ability to predict future preferences (Bélanger et al. 2014; Lee & Atance 2016) and delay gratification (Prencipe & Zelazo 2005). Social distancing also improves executive functioning in this same age group. White and Carlson (2016) found that preschoolers who took the perspective of another character (e.g., Batman) during an executive functioning task showed increased performance compared to preschoolers whose perspective was focused on the self. This effect was, however, more pronounced for 5- than 3-year-olds suggesting that even though young preschoolers are sensitive to social distance manipulations, this sensitivity increases with age.

Interestingly, however, preliminary work also suggests that children's sensitivity to spatial distance emerges later than sensitivity to social distance. For example, although taking a socially distanced perspective improved preschoolers' ability to predict future preferences (Bélanger et al. 2014) and delay gratification (Prencipe & Zelazo 2005), Rutt et al. (2019) recently showed that a spatial distance manipulation had no such facilitative effect. Specifically, 3- to 6-year-olds who were asked to take a spatially distanced perspective by imagining themselves in a far-away place were no more accurate in predicting future preferences, nor delaying gratification, than children who adopted a spatially near perspective (i.e., their current location). Similarly, Bowman-Smith et al. (2019) found that a spatial distance manipulation did not have a significant effect on children's reasoning until 6½. These findings, thus, suggest important differences between mental travel across social and spatial distances.

Findings by Coughlin et al. (2019) provide further evidence of differences between forms of mental travel. They compared 5-year-olds', 11-year-olds', and adults' ability to produce "temporal" and "make-believe" (i.e., hypothetical mental travel) narratives about a particular event (e.g., eating something yummy). Five- to 11-year-olds required more prompts to successfully produce future event narratives, and also received lower episodicity (i.e., amount of episodic detail) scores on future narratives, as compared to make-believe event narratives. Older children also required fewer prompts than younger children to produce future event narratives, whereas the number of prompts required for make-believe events did not differ with age. These findings suggest that the ability to mentally traverse time emerges later than the ability to mentally traverse hypotheticality.

Therefore, although we agree that different forms of mental travel might be largely interchangeable in adults, they appear to develop in a staggered manner in childhood. Though more data are needed, traversing (or being sensitive to) social (Bélanger et al. 2014; Lee & Atance 2016) and hypothetical distances (Coughlin et al. 2019) appear to emerge earlier in development than traversing temporal (Coughlin et al. 2019) and spatial distances (Bowman-Smith et al. 2019; Rutt et al. 2019).

In light of this, it is unsurprising that the mental representations (or forms of "abstraction") subserving the various dimensions of mental travel also appear to differ. For example, as described earlier, Coughlin et al. (2019) found that young children were better able to episodically pre-experience a hypothetical, than a future, event even though both dimensions require event simulation. Importantly, these authors found that self-concept coherence (i.e., how coherently an individual views him or herself) predicted future, but not make-believe, event narrative generation in younger children. Therefore, the cognitive concept of the self might be required to engage in mental travel across the temporal, but not the hypothetical, dimension. Given that self-concept coherence improves significantly with age (Coughlin et al. 2019), it makes sense that traversing the temporal dimension emerges later than traversing the hypothetical one. Similarly, contrary to what might be expected were all dimensions equal, Hanson et al. (2014) found that 3- to 5-year-olds' performance on theory of mind tasks (i.e., traversing social distance) was not related to their performance on episodic foresight tasks (i.e., traversing temporal distance). In sum, these are not the findings we would expect based on the idea that the capacity to mentally travel across different dimensions co-occurs in ontogeny.

Future research should systematically test children's sensitivity to the various dimensions in a single task (e.g., producing event narratives) while varying the dimension that must be traversed. For example, based on our account, children may show greater ease imagining how an event might unfold for another child than they do imagining how this event may unfold at a distant spatial location, or even for their future selves. This may be why in certain contexts, children's future-oriented decision-making is more accurate and adaptive for "other" than "self" (Bélanger et al. 2014; Prencipe & Zelazo 2005; for a similar study with adults, see Renoult et al. 2016). Further exploring the cognitive correlates of each form of mental travel would also continue to shed light on why some forms may emerge earlier than others. The results of such efforts may not fundamentally contradict Gilead et al.'s claims, but may instead add richness and precision to an account of how humans develop the remarkable capacity for mental travel.

Abstraction still holds its feet on the ground

Mariella Pazzaglia[a,b] and Erik Leemhuis[a,b]

[a]Department of Psychology, University of Rome "La Sapienza," 00185 Rome, Italy and [b]IRCCS Santa Lucia Foundation, 00100 Rome, Italy.
mariella.pazzaglia@uniroma1.it
http://dippsy.psi.uniroma1.it/users/pazzaglia-mariella
erik.leemhuis@uniroma1.it

doi:10.1017/S0140525X19003224, e141

Abstract

In view of current scientific knowledge, it seems premature to hypothesize a qualitative distinction between processes, networks, and structures involved in abstract processes from those based on perception, episodic, or procedural memories. Predictive thought and mental travel strongly rely, at different levels of consciousness, on past and ongoing sensory input, bodily information (e.g., interoception), and the results of perceptual elaboration.

Understanding the constituent elements, process organization, and neural networks involved in abstraction and predictive thinking requires confronting an extensive range of cognitive processes. In proposing their model, the authors have additionally integrated the different perspectives based on the principle of parsimony, essential to reductionism in research. However, some aspects, though deserving of more contextual prominence, have been relegated to the background.

Although we do not intend to discuss the role of language in complex abstract thought, it is the only way to access complex abstract; therefore, similar to other aspects of cognition, language is intertwined with sensorimotor and social information. Thus, question on the "hierarchy of abstractness" is extended to single words and their meaning but the concept of "abstract elements" necessitates a definition, and researchers occasionally have varied hierarchical vision. Gilead and colleagues hypothesized a discontinuity at some point between concrete and intangible in their abstraction hierarchy. Studies on explicit word categorization suggest the abstract as more nuanced than a simple difference between perceptible or intangible. Troche et al. (2017) hypothesized a continuous three-dimensional space (endogenous, exogenous, and magnitude) wherein the elements are placed according to their level of concreteness. A similar semantic classification study refuted the dichotomic view of concreteness–abstractness (Borghi et al. 2019b). Both abstract and concrete concepts involve the sensorimotor system and involve many dimensions, including social, linguistic, and inner experience.

Although an explicit dichotomous or mono-dimensional classification of concreteness–abstractness appears impossible, a diffused and inextricable interconnection seems viable even at the level of cortical representation of semantic contents. Another fascinating study used fMRI to draw a cortical semantic representation map (Huth et al. 2016). The findings revealed a similar organization between subjects, consistent with that based on semantic domains (i.e., abstract, social, and perception-related words), merging one into the other intricately. This organization complicates the visualization of clear boundaries between abstract concepts and episodic or sensorimotor memories, or other concurrent perceptive activity.

On a wider perspective, Gilead and colleagues offered another point on evolutionary interpretation. It cannot be denied that the products of the human mind extend beyond sensory stimuli-mediated concreteness and environmental perception. Simultaneously, it is widely accepted that the senses and complex behavior (e.g., environmental adaptation and mating strategies) are strictly interdependent. In our opinion, this interaction has profoundly directed evolutionary processes, including that in humans, making it difficult to hypothesize the mental processes, abstract contents, and neural substrates underlying them, and which are independent of perception.

Although the authors admit the multi-modal nature of perceptual experiences, differences in perceptual, interoceptive and motor pathways require additional emphasis, mainly in predictive contexts. As a result, the proposed model must account for abstract concepts in terms of two aspects that are crucial for cognitive, affective, and social processes. First, strictly body-generated information, such as proprioception and interoception, is not considered in the target article even though abstract words can be directly experienced through the homeostatic condition of the body (e.g., *hungry* and *fear*). Moreover, interoception and other bodily signals ground emotional concepts and should be thought as part of the system of abstract concepts (Connell et al. 2018). Second, one must consider the presence of an organization more complex and rooted than all perception-based information of telereceptive senses. The multimodality of perceptual experiences appears more effective with the inclusion of weak and sparse stimuli, despite their spatiotemporal co-occurrence (Ghazanfar & Lemus 2010). Reciprocal interactions between sensory domains prevent multisensory collaboration from being democratic or steady.

Studies on cross-modal perception indicate the circumstantial predominance of specific perception over multisensory elaboration. Action perception in routine life, for example, typically implies the merging of discrete information from the visual, auditory, or olfactory senses to optimally guide interpretation and behavior (Chen & Spence 2010). Although multiple senses mediate our experience with the world, sight usually provides the most reliable information for human predictive cognition. Vision is also the medium for studying the social and emotional nature of experiences (e.g., pain and anger) and for shaping, implementing, and adjusting action plans (Aglioti & Pazzaglia 2011). The predominance of vision over the other senses questions whether multisensory integration is necessarily mediated by visual transformations.

However, auditory and olfactory stimuli associated with abstract symbols, filter through multisensory redundant cues, occasionally dominate one's knowledge. Odors can not only influence emotional states and social interactions, modulating words and facial recognition, but also become associated with abstract symbols, influencing subsequent elaboration processes (Seo et al. 2010), such as anticipatory action planning (Aglioti & Pazzaglia 2010). Therefore, the sensory consequences of an odor are integrated with the ongoing high-order cognitive activity. From a neural view point, the olfactory networks still remain largely unexplored despite strong evidence of their involvement in the activation of a network of brain areas typically responsible for high-order processes (Zhou et al. 2019). Similarly, the auditory system presents characteristics that predominate other senses (Aglioti & Pazzaglia 2010), with sounds helping to establish a state of alertness, predict upcoming events, and trigger anticipatory representations (Aglioti & Pazzaglia 2011). The brain areas identified in these studies include those involved in sensory and motor processing, presenting compelling evidence in favor of modality-specific abstract information. Furthermore, the elaboration of sensory information is a continuous and implicit process that may never reach a conscious level of perception but influence ongoing high-order cognitive activity (Walla 2008). Although modality-specific sensory information is important in forming the perceptual features of concrete representation, sensory afferences deeply influence the higher cognitive processes that characterize real-world experiences.

Thus, it is suggested that unimodal input for prediction provides more rapid and less variable acquisition, more precise dynamic representations, and more specific information for anticipatory coding. From an evolutionary view point, knowledge of forthcoming events is more significant than past events. Analyzing the contribution of these different sensory channels and bodily information to the perspective coding of abstract and concrete concepts, their influence, or role in defining the content and form of mental travel still remains a fundamental topic for future research.

Neuronal codes for predictive processing in cortical layers

Lucy S. Petro[a] and Lars Muckli[b]

[a]Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB, UK and [b]Centre for Cognitive Neuroimaging, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB, UK.
Lucy.Petro@glasgow.ac.uk
Lars.Muckli@glasgow.ac.uk
https://muckli.psy.gla.ac.uk/

doi:10.1017/S0140525X19003121, e142

Abstract

Predictive processing as a computational motif of the neocortex needs to be elaborated into theories of higher cognitive functions that include simulating future behavioural outcomes. We contribute to the neuroscientific perspective of predictive processing as a foundation for the proposed representational architectures of the mind.

By endeavouring to integrate predictive processing theories of brain function with human capacity for prospective thought, Gilead and colleagues have identified a subject matter that invites rigorous empirical focus. The current evidence for predictive processing is biased towards the field of perception. Such data are essential, but this bias leaves open questions about how neuronal prediction contributes to "offline" brain processing in which we can flexibly traverse space and time in our mental activity. As cognition is predictive, we should now pursue a framework for predictive neuronal processing that is 'abstracted away' from proximal sensory inputs and used instead for prospective deduction or future simulation. The authors have initiated a thought-provoking account of how abstract representational elements supporting predictive cognition could be structured, and their structure bridges from perception science to cognitive psychology. We highlight the neurobiology of predictive processing as a subdivision for developing this framework.

Predictive brain frameworks prescribe specific neuronal processes. These processes can be summarised such that the hierarchical brain compares sensory inputs with internally generated models of the world, aiming to minimise error signals that indicate a mismatch and result in internal model revision (Friston 2005). There are various mechanistic implementations of this computation (George & Hawkins 2009; Spratling 2017), but broadly speaking, neuronal markers of prediction should reveal (aspects of) these coding principles. The authors have examined how neuroscience data, specifically human functional neuroimaging, corroborate their proposed structure of abstract representations. Brain signatures have been observed that support the distinct abstracta they describe, especially so for modality-specific and multimodal abstracta. The element of the model where 'abstractness' is greatest, that is categorical abstracta and predicator representation, should be investigated more sufficiently in brain imaging experiments testing predictive processing, but the authors provide a plausible hypothesis for how this could be realised in specific hubs of the default mode network (DMN). Under the assumption that identical computations (i.e., prediction) support all functions throughout the brain's hierarchy, the distinction between perceptual and cognitive processing is eliminated. As such, future data should confirm the hierarchy of inference differs in the content of mental representation at each level (not the computation), and areas processing abstract representations for predictive cognition should reveal neuronal indicators of prediction. In line with this, the putative involvement of the DMN in 'offline' processing to probabilistically simulate future outcomes can be described in the context of Markov decision processes (Dohmatob et al. 2018). This mathematical framework is in line with predictive neuronal processing in the DMN, which is supported by experiments showing that the DMN overlaps with areas involved in forming associations (from which predictions are derived, Bar 2007).

How close are we to testing a model of the predictive neural bases of conceptual cognition? Taking the perspective that the brain performs predictive cognition for mental time travel, one approach is to test the brain's generation of sensory inputs that are decoupled from perception. Mental imagery can be used to test internally-driven events, and such data have been interpreted within predictive processing frameworks. For example, analysis of functional magnetic resonance imaging data reveals that vividness of visual imagery increases the top-down coupling of signals from frontal areas to early visual areas (Dijkstra et al. 2017). This finding fits with the role of cortical feedback carrying descending predictions from high to low areas, even in the absence of sensory input. Further, in primary sensory areas, brain imaging studies of illusions and mental imagery reveal that processing during perception is generative, for example, as observed in visual counterstreams. In primary visual cortex it is possible to partition feedforward and feedback signals; this is crucial because these pathways transmit sensory versus internal processes respectively (or ascending prediction errors and descending predictions). We propose studying cortical layers in humans as a neuronal substrate allowing for mental time travel alongside perceptual processing. We motivate this hypothesis using high-resolution brain imaging showing that neuronal codes exist for abstract mental representations transferred by cortical feedback to sensory areas (Bergmann et al. 2019), and that sensory cortex allows for predictive processing mechanisms to be precisely spatiotemporally mapped (e.g., Edwards et al. 2017). Taking visual imagery as an example of counterfactual processing, deep cortical layers of visual cortex are involved in maintaining visual information specific to mental imagery (Bergmann et al. 2019). A dual stream of factual and counterfactual information processing might provide insight to higher cognitive functions in which potential future outcomes need to be simulated before informed decisions can be executed. The inherent challenge however in investigating prediction machinery in higher, abstract, psychological function is that we have diminished control over the content of internal representations, and we have to probe brain areas that do not easily allow for the separation of sensory from internally-generated processes.

Developing the authors' account will now require, in part, overcoming the challenge of understanding if and how neuronal substrates of abstract mental representation, even those that confer predictive mental abilities such as projecting oneself forward in time, are one and the same as neuronal prediction. Towards the aim of a testable model of the neural bases of conceptual cognition, it will be essential to develop the description of higher cognitive functions within the parsimonious framework of predictive processing. This framework should span disciplines by defining common concepts and terminology. In doing so, it is imperative to avoid a language surplus as seen previously with terminology that was justified either on behavioural data alone, or in connection with specific paradigms, without providing a reduction into descriptions of neuronal processes. For example, attention has often been defined based on behavioural advantages without direct neuronal explanations, but might be more thoroughly and accurately explained as optimising the precision of prediction error by increasing the synaptic gain of these neurons (Feldman and Friston 2010; Gordon et al. 2019). The target article provides a descriptive framework of abstract representations for predictive cognition onto which we can map neuronal data, which should comprise neuronal recordings, simulated data and models and behavioural measures in order to advance a parsimonious conceptual framework.

Dynamic hierarchical cognition: Music and language demand further types of *abstracta*

Tudor Popescu and W. Tecumseh Fitch

Department of Cognitive Biology, University of Vienna, 1090 Vienna, Austria.
tudor.popescu@univie.ac.at
https://cogbio.univie.ac.at/people/postdoctoral-researchers/tudor-popescu/
tecumseh.fitch@univie.ac.at
https://cogbio.univie.ac.at/people/staff/tecumseh-fitch/

doi:10.1017/S0140525X1900308X, e143

Abstract

Hierarchical structures are rapidly and flexibly built up in the domains of human language and music. These domains require a tree-building capacity – "**dendrophilia**" – to dynamically infer hierarchical structures from sensory input (or to hierarchically structure output), based on subunits stored in a lexicon. This dynamic process involves a crucial class of *abstracta* overlooked in the target article.

Gilead et al. wisely avoid assuming that abstraction and mental time-travel are exclusively confined to human cognition (contra Penn et al. 2008; Suddendorf and Corballis 2007; 2010), because ample evidence exists for such abilities in animals (e.g., Emery and Clayton 2004). However, the propensity to perceive hierarchical structure in sequences unfolding through time or space, in multiple modalities, does seem unusually well-developed in our species. This propensity to generate tree-like structures has been dubbed "dendrophilia" (Fitch 2014).

We share the authors' conviction that prediction is fundamental to cognition, and applaud their pluralistic perspective regarding the many abstraction types that support prediction. However, their ontology of abstraction overlooks the importance of the dynamically constructed hierarchies fundamental to music and language. What enables these domains to make "infinite use of finite means" is precisely their hierarchical combinatoriality. Despite pervasive evidence for hierarchical structures in human cognition, evidence for hierarchical representations in nonhuman animals remains sparse (e.g., Jiang et al. 2018; Wang et al. 2015).

Music and language obviously involve hierarchical structures, whereby, following Gilead et al., *abstracta* at one level of abstraction act as the *concreta* for another. Thus, music's *concreta* include individual **notes** at certain pitches, whereas speech builds on individual **phones**. At the first level, the *abstracta* are **pitch classes** (the set of pitches that share the same chroma across different octaves, e.g., C_3, C_4, etc.) and **phonemes** (incorporating different allophones). Further hierarchical structures compose musical *abstracta* into **melodies** (sequences of notes of different pitch classes with a particular rhythmic structure) and musical **keys** (subsets of certain note classes) whereas language composes phonemes into syllables, morphemes, phrases, sentences, and narratives.

Importantly, the "terminals" in both music and language are *themselves* abstracta. For instance, a phoneme is not a concrete entity that exists in nature, but is already processed by categorical perception, thus is *already* an abstraction. Similarly, the "same" melody can be produced with any arbitrary starting note (and thus encompasses an infinite set of surface frequency sequences).

These static structures are consistent with the kind of static (taxonomic) hierarchical structures that Gilead and colleagues consider (e.g., dogs are mammals are vertebrates). But, these do not account for the type of flexible generativity needed to create or understand new sentences or melodies, which requires smaller stored subunits (words, idioms, or short chord sequences) that can flexibly combine into larger wholes. Such compositional subunits go by many names ("subtrees," "treelets," and "curried functions") and are a generalisation of the notion of "predicate," whose importance the authors rightly emphasise.

Extending Frege's terminology, such "partially saturated entities" are the crucial building blocks of the larger hierarchical structures underlying music and language. But in music, these subunits lack compositional semantics: unlike predicates they do not imply truth values. They are purely combinatoric entities, subunits with open "slots" that allow them to flexibly combine with other subunits in constrained ways. Thus, just as a transitive verb has "slots" demanding an animate subject and an object, a tonic-subdominant-dominant chord sequence allows a constrained set of continuations (canonically a return to the tonic, but unlimited variants are possible). Such compositional chord sub-sequences are not predicates: they make no semantic reference to the world, and entail no truth values.

Crucially, the hierarchical subunits of music and language are learned, and must be stored in a mental lexicon, but are deployed automatically and habitually. Thus, during ontogeny, such learned subunits must be **automatised**. This blurs the line between the traditional category of habitual or "model-free" abstracta and (conscious) model-based prediction. That some predictions are more habitual or unconscious than others does not mean they are "model-free": only that the abstracta involved have been automatised.

Consider how we perceive music, where the crucial first level of abstraction is constituted by traversing *temporal* distance. Musical abstraction applies to the relatively short time-spans of unfolding chords or melodic phrases. These units nest into larger ones, that can be described at increasing levels of abstraction, in terms of for instance **Schenkerian time-span reduction**, whereby each phrase can be conceived of as an elaboration of a basic underlying tonic-dominant-tonic progression (Schenker 1935). Such dynamic hierarchical abstractions are reflected in electrophysiological evidence for the ability to recognise long-distance harmonic dependencies in music (Koelsch et al. 2013). Our unconscious knowledge of music is essentially a **generative model** for an infinite set of melodies (Lerdahl & Jackendoff 1983), built upon a finite set of hierarchical building blocks that combine to form more complex hierarchical structures.

This parallels language, where the learned structures stored in the lexicon (including both morphemes and more complex subunits like idioms; cf. Jackendoff 2002), are used online to build up the more complex and variable structures involved in language use. Available data from psycholinguistics indicate that complex structures are built up "on the fly," in parallel, and unconsciously (Cutler 2017). These processes are not easily available to introspection because they have been automatised. Both cases illustrate the centrality of dynamically created hierarchical structures in human cognition, and the musical case demonstrates that semantics (and thus predicates) need not be involved.

Consider finally the aesthetic dimension of music or poetry. The temporal pattern in which tension unfolds over time, where expectations are unconsciously generated, then challenged, and finally resolved and integrated into a satisfying whole, can be termed an "**aesthetic trajectory**" (Fitch et al. 2009). In music, the ebb and flow of the aesthetic trajectory operates at the "top" level of this hierarchy, that is, on melodies and keys, not on individual notes. Such dynamic aesthetic trajectories are fundamental to aesthetic experience (including, we argue, to static arts like painting or sculpture; cf. Fitch et al. 2009), but also far beyond, for example, in following narratives, getting jokes, or understanding pragmatics (Hurley et al. 2011).

We conclude that a dendrophilic process of dynamic hierarchical structure building is crucial in music and language, and deserves a prominent place in Gilead's pluralistic ontology of abstracta. This process is based upon combinations of automatised subunits, analogous to but more general than Fregean predicates.

Abstraction: An alternative neurocognitive account of recognition, prediction, and decision making

Valerie F. Reyna[a] and David A. Broniatowski[b]

[a]Department of Human Development, Human Neuroscience Institute, Cornell University, Ithaca, NY 14853 and [b]Department of Engineering Management and Systems Engineering, The George Washington University, Washington, DC 20052.
vr53@cornell.edu
https://www.human.cornell.edu/hd/research/labs/lrdm/publications
broniatowski@gwu.edu https://www2.seas.gwu.edu/~broniatowski/index.html

doi:10.1017/S0140525X19003017, e144

Abstract

Gilead et al. offer a thoughtful and much-needed treatment of abstraction. However, it fails to build on an extensive literature on *abstraction*, *representational diversity*, *neurocognition*, and *psychopathology* that provides important constraints and alternative evidence-based conceptions. We draw on conceptions in software engineering, socio-technical systems engineering, and a neurocognitive theory with abstract representations of gist at its core, fuzzy-trace theory.

We commend the authors for a thoughtful and much-needed focus on abstraction, integrating developmental, cognitive, social, and neuroscience literature. This discussion and future work would benefit from drawing on perspectives from software and systems engineering incorporated into recent formalizations of fuzzy-trace theory (Broniatowski & Reyna 2018). Fuzzy-trace theory has spawned an extensive literature that directly relates to *abstraction*, *representational diversity*, *neurocognition*, and *psychopathology* that provides important constraints and alternative evidence-based conceptions of abstraction.

Specifically, the theory distinguishes verbatim representation of information – symbolic representation of concrete, surface form – from gist representation – "fuzzy" bottom-line meaning (Reyna 2012). Gist varies in abstraction of content – called "hierarchy of gist." In socio-technical systems engineering, Rasmussen (1985) defines two types of hierarchies: abstraction and aggregation. Abstraction pertains to gist in that it emphasizes embedding items within their contexts such that they have a meaningful purpose. Aggregation emphasizes combination of parts into wholes such that "higher" elements in the hierarchy contain "lower" elements, hiding their contents.

We apply both kinds of hierarchies. Implementing fuzzy-trace theory, we define mental representations that vary in abstraction – categorical gist, ordinal gist, and interval verbatim. However, gist-based abstraction is not verbatim-based aggregation, and gist is not derived from verbatim representations. Aggregation rules can take verbatim representations as input, such as combining probabilities and outcomes using *rote* procedures into precise expected values, but this is not abstraction in the gist sense. Our literature review contradicts the assertion that "every action an organism makes is an attempt to reduce uncertainty"; decision makers sometimes seek uncertainty (e.g., for losses), which has profound implications for legal (rejecting plea bargains) and medical decision-making (seeking experimental treatment when terminally ill).

Notably, more abstract levels of the gist hierarchy use a looser, more parsimonious rule (compared to other decision theories) by specifying a partial order in which representations can be, but need not be, ordered. This looser rule reflects fuzzy-trace theory's emphasis on fuzzy processing. Broadly, within each level of representation, different aggregation procedures take the form of different order relations between elements in each set. Thus, we integrate notions of abstraction found in software engineering and socio-technical systems engineering with fuzzy-trace theory's levels of mental representation.

These theoretical ideas predict numerous results that are directly relevant to the authors' claims. Space limitations permit only a few illustrations. For example, with its roots in psycholinguistics, fuzzy-trace theory absorbed concepts such as scripts/schemas cited by the authors, but modified them to accommodate serious empirical contradictions (e.g., Alba & Hasher 1983). The theory also goes beyond dualisms such as model-free versus model-based or declarative versus procedural memory because data demanded it. The old ideas that "Although episodic memory is more detailed and concrete than semantic memory, it is nonetheless *declarative*; namely, information that can be readily put into words" and "as such may be considered more abstract" have been superseded by evidence showing that episodic memory consists of both verbatim and gist memories. The whole of the false-memory literature speaks against the notion that such memories "can be readily put into words." Moreover, the ability to articulate cognitive processes, such as memory, is a faculty distinct from representational abstraction (as much research showed). In any case, words are not "considered more abstract," but, rather, vary in abstraction. Phenomenology of a false memory (vivid) should not be confused with the abstractness of the representation: "concrete, vivid simulations can easily give rise to false memories." The opposite is true; concreteness and imageability are *negatively* related to false memory, per theory (Brainerd et al. 2008).

Further, prediction does not necessarily entail abstraction (or deliberation). Indeed, current reinforcement and prediction-error paradigms, especially in cognitive neuroscience, emphasize rote memorization. However, in fuzzy-trace theory, there are two main ways people make predictions: through verbatim memory for event frequencies (the number of times an event has happened) or gist-based beliefs about why events occur (Mills et al. 2008). Neurocognitive underpinnings for this fuzzy-trace-theory distinction have been delineated (d'Acremont et al. 2013; see Spreng & Turner 2019, for a theoretically compatible neurocognitive theory that shares the authors' focus on the default-mode network). Thus, rote memory for frequency counts is not abstract, it reflects concrete experience, as does mindless association of events solely through "spatiotemporal contiguity." Although association between events can be abstract in the mind of the theorist, it is not necessarily abstract in the mind of a rat. Any two things that can be related are not necessarily related abstractly.

When abstraction was manipulated per construal level theory, its effect on decision quality was mediated by gist representations (a potential rapprochement; Fukukura et al. 2013). Contrary to other theories, fuzzier, less detailed representations are associated with contextually biased but higher-quality developmentally *advanced* decision-making (Helm et al. 2018; Reyna et al. 2011). Extensive evidence contradicts the assertion that "In order for a target-representation to be functional, it must be accurate *and* detailed. When either condition is not met, the target representation is useless."

According to fuzzy-trace theory, representational diversity, then, is realized whenever people encode information, but the relative emphasis on different levels of abstraction varies across age and individual differences (Reyna & Brainerd 2011). Among those individual differences, brain and behavioral analyses have linked (lack of) abstract thinking to "psychopathology," what we describe as atypical information-processing, including autism, adult (as opposed to adolescent) criminality, and psychopathy (Reyna & Panagiotopoulos, in press; Reyna et al. 2018). For example, noncriminal risk-taking behavior was associated with emotional reactivity (amygdala) and reward motivation (striatal) areas, whereas criminal behavior was associated with greater activation in temporal and parietal cortices, their junction, and insula. Neurocognitive and experimental evidence converged on the conclusion that psychopathology was associated with more objective, seemingly rational verbatim processing, rather than developmentally typical reliance on abstract (but contextually biased) gist in adulthood.

In sum, tasks, such as recognition or prospection, are solved differently using concrete verbatim-based processing and abstract gist-based processing. This distinction predicts some effects discussed by the authors and fundamentally contradicts others. Scientific progress can best be achieved by integrating prior evidence with the authors' exciting ideas about abstraction.

Shared reality and abstraction: The social nature of predictive models

Maya Rossignac-Milon[a], Federica Pinelli[b] and E. Tory Higgins[c]

[a]Management Division, Columbia Business School, Columbia University, New York, NY 10027; [b]Psychology Department, Columbia University, New York, NY 10027 and [c]Psychology Department, Columbia University, New York, NY 10027.
mr3352@columbia.edu
eth1@columbia.edu
fp2224@columbia.edu

doi:10.1017/S0140525X19003212, e145

Abstract

We propose that abstraction is an interpersonal process and serves a social function. Research on shared reality shows that in communication, people raise their level of abstraction in order to create a common understanding with their communication partner, which can subsequently distort their mental representation of the object of communication. This work demonstrates that, beyond building accurate models, abstraction also functions to build *socially shared* models – to create a shared reality.

In their paper, Gilead et al. make an important contribution to the field of predictive cognition by leveraging construal level theory (CLT; Trope & Liberman 2010) as a unifying framework underlying various forms of abstract representation. In this commentary, we situate their contribution in terms of the interpersonal context. We propose that predictive models are not generated and used solely within the confines of the individual mind – instead, these are fundamentally socially constructed. We examine abstraction as an interpersonal process and argue that it serves a social function.

As Gilead and colleagues point out, the minds of other people may be highly relevant to the predictive process in that people "wish to align themselves with the beliefs of others" (sect. 3, para. 7), even for non-social stimuli, and that "contrary to theories that suggest that prediction is purely an epistemic process, we highlight that in humans, mental travel is often guided by the motivation to arrive at a state of shared belief with other humans – regardless of the truthfulness of those beliefs" (sect. 3.3, para. 1). We strongly agree with this position (Echterhoff et al. 2009; Hardin & Higgins 1996; Higgins 2019; Rossignac-Milon & Higgins 2018a). But further, we propose that predictive models are socially constructed more often than not. Evidence from the field of shared reality supports this idea.

Shared reality theory proposes that humans are fundamentally motivated to turn to each other in order to validate their interpretations of the world and understand what is real and true (Echterhoff & Higgins 2017; Hardin & Higgins 1996; Higgins 2019). Research on the Sharing-Is-Believing effect (Higgins 1992; for reviews, see Echterhoff & Higgins, in press; Higgins 2019) has shown that in communication, people tune what they say to fit with their communication partner's perspective on the object of communication (i.e., the target referent). For example, when describing a target person's behaviors to a colleague, people will categorize (i.e., abstract) the behaviors more positively (vs. negatively) if they know that their colleague likes (vs. dislikes) this person. More importantly, when speakers feel connected to their audience, they subsequently align their cognitive representations of the target to fit with the evaluative tone of what they said. Thus, they integrate their communication partner's attitude into their abstract mental representation of the target.

As described by CLT, attitudes are judgments about the "essence" of an object, which are more abstract than its concrete details (Trope & Liberman 2010). Critically, the new evaluatively-tuned mental representation is a *distortion* of the original given information about the target's behaviors. Thus, the attitudes of others can significantly shape abstraction, even at the cost of accuracy. Indeed, Gilead and colleagues touch on the uniquely human ability to disregard different dimensions in order to create *intangible* abstracta. These intangible abstracta are created not only to summarize information, but also (and above all) in the service of creating a common representation – a shared reality. As demonstrated in the Sharing-Is-Believing paradigm, the communicator produces an evaluatively biased message based on the audience's attitude, in the service of creating a shared abstractum that reflects their common understanding of the object. Moreover, the communicator's message can also bias the listener's understanding (Hirst & Coman 2018), thereby further aligning their understanding.

As suggested by Gilead and colleagues, categorical representations are primarily instantiated through social interaction. Categorization and multimodal abstracta serve a communicative purpose: when determining which level of abstractness to use (and specifically, which lemma; Roelofs 1992), people take into account the perspective of their conversation partner and select the one most useful for the conversation at hand; for example, people refer to spare change as "coins" and not "round metal objects" because their monetary utility is most relevant to their conversation partner (cf. Brown 1958). Communication goals determine the selected level of abstractness.

In fact, people may raise the level of abstractness in order to create shared realities. Indeed, Gilead and colleagues assert that the function of beliefs is not always to "accurately represent reality – but rather to facilitate traversing social distance by creating unity of minds" (sect. 3, para. 8). We agree with this point. The purpose of abstraction is not simply to minimize prediction error. Predictive models function to minimize the discrepancy not only between one's internal state and the state of the world, but also between one's internal state and *that of another person*. Instead of accuracy, the motivation can be to create something that "we" believe in *together* (Rossignac-Milon & Higgins 2018b). Although this "we" is an abstract category that subsumes "me" and "you" – thereby *increasing psychological distance* away from the "me here and now" (Yip-Bannicq 2018) – it also *decreases social distance*. As CLT would predict, this decreased social distance should simultaneously collapse perceived spatial and temporal distance. The social glue created by shared reality was critical to our survival as a species in our ancestral environment (Higgins 2019). Thus, we argue that the evolutionary pressure to raise abstractness stems not only from accurately predicting the environment, but also from the need to create shared realities with others.

Moreover, through this process of creating shared realities, *conversation partners jointly align their predictive models* (e.g., Hirst & Echterhoff 2012; for a review, see Higgins 2019). As mentioned by Gilead and colleagues, predictive models can be transmitted from person to person (e.g., parent to child). Another understudied avenue through which predictive models can become shared is through dynamic, dyadic alignment. In this *co-construction* process, a model is not transmitted from one mind to the other; instead, it is an *emergent product of both minds*. We believe that these co-constructed predictive models have the greatest psychological weight (see Rossignac-Milon and Higgins 2018a).

In conclusion, the field of predictive cognition disproportionately emphasizes intra-personal processes. Humans do not build predictive models of the environment in social isolation. People verify their predictive models not only by determining whether they can predict subsequent events, but also by determining whether other people agree with them. Abstraction functions to build not only accurate, but also *socially shared* models.

Note. There was an editorial error in the abstract of the original online version of this commentary. It has been corrected here and an erratum has been published.

Acknowledgment. This work was supported by the Anneliese Maier Research Award from the Alexander von Humboldt Foundation awarded to E. T. Higgins.

Abstractions, predictions, and speech sound representations

Mathias Scharinger[a,b]

[a]Phonetics Research Group, Department of German Linguistics, Philipps-University, 35032 Marburg, Germany and [b]Center for Mind, Brain & Behavior, 35032 Marburg Germany.
mathias.scharinger@gmail.com http://www.inter-word.net/home/

doi:10.1017/S0140525X19003030, e146

Abstract

Gilead et al. provide a unified account of predictive cognition in which abstract representations play an essential role. Although acknowledging the similarity to linguistic concepts toward the higher end of the proposed abstraction gradient, Gilead et al. do not consider the potential of their account to embrace phonetic and phonological speech sound representations and their neural bases.

Within linguistics and the cognitive sciences, some theories rely heavily on representations (Markman & Dietrich 2000). As generic constructs, representations may be characterized as memory traces (Rosen 1975), categories (Rosch 1973), or concepts (Medin 1989), depending on the level of abstraction pursued in the respective frameworks. Common to most approaches is the assumption that representations "stand for entities" in the environment (Ramsey 2010), that is, that they provide essential units for simulating the external world. As such, representations are readily suited for predictive coding frameworks (Gładziejewski 2015) that postulate that cognition is based on simulation, active inference, and hypothesis testing (Friston 2005). In this vein, Gilead et al. propose that predictive cognition necessitates representations that are hierarchically organized along an abstraction continuum. The advantage of their approach is that it operationalizes the definition of abstraction, is readily applicable to phonetic and phonological speech sound representations, and is compatible with a range of neurolinguistic findings.

Several theories on sound representations in human language propose different levels of abstraction. Rooted in generative linguistics, a basic distinction is made between (more abstract) underlying representations and (less abstract) surface representations (Chomsky & Halle 1968). This distinction can also be aligned with the distinction between phonology and phonetics. A phonological representation, in these approaches, is a phonemic (or meaning-distinguishing) representation of a sound in minimal pairs such as *sum* and *sun*. A phonetic representation, on the other hand, may be more detailed and provide additional information about pronunciation variants of the two word-final nasal consonants [m] and [n]. Abstraction is operationalized in approaches that consider speech sound representations to be a bundle of phonological or phonetic features (Clements 1985; Jakobson et al. 1965; Stevens 2002). The notion of abstractness in these theories closely resembles the definition provided by Gilead et al., in that abstractness means "two or more subjectively distinguishable objects [that] satisfy" the structural descriptions of the proposed representation (sect. 1, para. 1, first item 3). Distinguishable objects in speech sounds refer to acoustic–phonetic dimensions such as the place of articulation of nasal consonants, distinguishing bilabial [m] (produced at the lips) from alveolar [n] (produced at the alveolar ridge behind the upper front teeth). In English, on a relatively high level of abstraction, it is assumed that the nasal in morphemic *in*– ("not," preceding *probable* or *tolerant*) shares the "subjectively [and phonetically] distinguishable objects" *bilabial* and *alveolar* place of articulation to allow for the variations seen in *improbable* and *intolerant*. Because both places of articulation cannot be expressed simultaneously, the assumption in phonological theories is that the abstract representation of nasal consonants does not include place of articulation information in such cases. This is referred to as *underspecification* (Lahiri & Reetz 2010; Steriade 1995): The nasal consonant in *improbable* and *intolerant* is said to be underspecified for place of articulation. In the process of speech sound production, the spell-out of more concrete representations (*concreta* in the terminology of Gilead et al.) is accomplished by phonological and phonetic rules (*criterion of substitutability* in the terminology of Gilead et al.). These rules insert the necessary place of articulation feature either as a result of a default setting or as a result of a contextual fill-in (Lahiri & Reetz 2010). Similar to Gilead et al.'s approach, the outputs of abstraction processes can serve as inputs for additional abstraction processes. The feature bundle describing [m] would contain the feature CONSONANTAL (distinguishing consonants from vowels), NASAL (distinguishing nasal and non-nasal consonants), and LABIAL (distinguishing labial and non-labial consonants). The additional act of abstraction in instantiating underspecification would then remove the feature LABIAL.

Abstract (underspecified) speech sound representations may relate to predictive processes in at least two ways. First, the spell-out of *concreta* (phonetic representations) can be context-specific in that following consonants determine the place of articulation of the preceding nasal consonant, leading to [m] when preceding bilabial [p] (*improbable*) and to [n] when preceding alveolar [t] (*intolerant*). Gilead et al. propose that "there is no mental travel without abstraction" (sect. 4, para. 2) where mental travel stands for *predicting future states*; in the speech example, predicting future states would mean predicting upcoming articulations and modifying implementations of abstract representations accordingly. During speech production, simulations of upcoming articulations account for the observation that upcoming speech sounds indeed exert influences on earlier speech sounds. This regressive assimilation has been related to anticipatory processing (Gow 2001). Internal simulations of future states may also minimize the prediction error in the production–perception loop involved in speech processing. Second, speech sounds differing in their assumed degree of abstraction seem to relate to differences in predictive processing. Broadly speaking, speech sounds based on more features allow for more precise or stronger predictions. Evidence for this observation stems from neurolinguistic studies (e.g., Scharinger et al. 2016) interpreted within an auditory predictive coding framework (Baldeweg 2006). Further evidence is provided by studies that suggest more concrete representations to be based on supplementary motor information (Möttönen & Watkins 2009; Möttönen et al. 2013). For instance, by disengaging the motor lip area, Möttönen et al. (2013) showed that representations of labial speech sounds (in *ba*-syllables) were disrupted, whereas representations of non-labial speech sounds (in *da*- or *ga*-syllables) were not.

Abstraction in speech sound representations as illustrated above is reflected in cortical structures in human temporal cortices where increasingly abstract speech sound representations are supported by regions with increasing distance to the primary auditory cortex (Humphries et al. 2014; Obleser & Eisner 2009). In these studies, more detailed and concrete (phonetic) representations were supported by regions in primary auditory cortex in the vicinity of Heschl's gyrus, whereas more abstract (phonological) representations recruited regions in more anterior and posterior parts of the superior temporal gyrus and sulcus. This cortical abstraction gradient is parallel to the one for visual perception by Gilead et al. and is in line with the aforementioned phonetic-phonology distinction.

In sum, Gilead et al.'s approach provides interesting starting points for future studies combining current theories of brain function with generative linguistic accounts of speech sound representations.

https://doi.org/10.1017/S0140525X19002000

Simulation across representation: The interplay of schemas and simulation-based inference on different levels of abstraction

Malte Schilling[a,b], Nancy Chang[c],
Katharina J. Rohlfing[d] and Michael Spranger[a]

[a]Sony Computer Science Laboratories Inc., 3-14-13, Higashigotanda, Shinagawa-ku, 141-0022 Tokyo, Japan; [b]Faculty of Technology, Department of Neuroinformatics, Bielefeld University, D-33501 Bielefeld, Germany; [c]Google, Mountain View, CA 94043; [d]Faculty of Arts and Humanities, Paderborn University, Paderborn, Germany.
mschilli@techfak.uni-bielefeld.de
ncchang@google.com
katharina.rohlfing@uni-paderborn.de
michael.spranger@gmail.com

doi:10.1017/S0140525X19003169, e147

Abstract

Language comprehension of action verbs recruits embodied representations in the brain that are assumed to invoke a mental simulation (e.g., "grasping a peanut"). This extends to abstract concepts, as well ("grasping an idea"). We, therefore, argue that mental simulation works across levels of abstractness and involves higher-level schematic structures that subsume a generic structure of actions and events.

Gilead and colleagues distinguish two specific reasoning mechanisms geared toward distinct levels of abstraction. First, they assume that a theory-based account is required for realizing inference on abstract levels. Second, they propose mental simulation as a means of extrapolating detailed imagination into possible futures. In contrast, we argue that mental simulation provides a flexible unified mechanism that supports a spectrum of (embodied) abstraction and inference driven by the schematic structure of action and event representations. Our proposal is based on a large body of neuroscientific and linguistic evidence arguing for the gradient nature of conceptual abstraction. The proposed type of simulation recruits (multi-)modal areas distributed across the brain as well as higher-level areas associated with more abstract and schematic concepts (Binder & Desai 2011).

Evidence of recruitment of brain areas in language use supports this view of mental simulation as involving flexible underlying representations and mechanisms across different levels of abstraction: action-related words recruit sensorimotor representations (Pulvermüller 2018). This process can be understood as grounding a mental simulation. Early experiments have established that this recruitment spans different levels of abstraction (Boulenger et al. 2008), a fact not mentioned by Gilead and colleagues. In a recent fMRI experiment, Wang et al. (2018) further showed how processing of abstract words recruits distinct high-level language-related areas, as well as widely distributed (multi-)modal brain regions. They conclude that abstract concepts recruit embodied representations but also connect to higher-level linguistic networks. Findings from Desai et al. (2013) support this view through analysis of brain activations for different levels of abstraction. In comparing activations for language related to concrete action, metaphorical meanings and idioms, they found differential activations of motor areas for action language and metaphors – suggesting an effect of mental simulation in both. However, idioms showed only graded activation in these areas, suggesting some degree of abstraction from sensorimotor systems. Desai et al. conclude that idioms undergo a process of conventionalization in which their representations shift from more concrete motor representations toward more schematic representations. These neuroscientific studies suggest that sensorimotor representations contribute directly to the formation of abstract concepts and are differentially recruited in mental simulations. Further evidence suggests their role in simulation can be modulated by grammatical features, for example, as found by Liu and Bergen (2016) in a study on how progressive and past tense affects language comprehension. They compared how speakers process sentences describing concrete versus abstract actions and found a location-sentence compatibility effect only for the progressive case. This result is consistent with the function of progressive tense as focusing on the internal details of the unfolding event, which requires a (more or less) detailed mental simulation. In contrast, the perfect tense shifts the focus from the means of action toward end-conditions after executing that action. Importantly, the effect is independent of using abstract or concrete verbs and reflects the shared schematic structure of actions.

We use the term "schematic structure" here to mean a categorical mental representation that generalizes over instances (Johnson 1987). Schemas capture regularities in sensorimotor systems that organize experience and interactions with the environment. In language comprehension, it is assumed that words and larger phrases activate schemas that add perceptual or motor components to a mental simulation of the content. These activations range from simple reinstantiations of detailed, modality-specific experiences up to abstract structures that capture basic conceptual knowledge. For instance, a source-path-goal schema is an abstract representation structuring our understanding of directed motion (Bergen & Chang 2005). Crucially, schemas are assumed to be "categorical" and rely on "predication" (as defined by the authors) across all levels of abstraction, because they allow for flexible recruitment in mental simulation across different levels of abstraction. Consider, for example, an action-schema for grasping: First, the use of parameterization (e.g., position where to grasp) is a form of abstraction. Second, a grasp-schema can subsume different types of grasping (pinch-grasp etc.) that share a generic event structure whereas involving different effectors or typical graspable-objects. Third, schemas can relate to each other (e.g., a grasp-schema requires an actor and a graspable object), implying abstraction over role-fillers. This view of schematic structure fulfills Gilead and colleagues' definition of abstraction as subsuming distinguishable schemas and actions. Moreover, the practical utility of such schemas for understanding abstract language has been nicely demonstrated by a computational model for metaphorical inference in the domain of economics (Narayanan 1999). In this account, the ability to parameterize mental simulations in the source domain served as the basis for further inference in the target domain.

A schematic event representation is as well supported by neuroscientific evidence: a gradient of abstraction for such representation is assumed along midline structures of the brain (Stawarczyk et al. 2019). This includes the default network and, in particular, posterior cingulate cortex as well as medial prefrontal cortex, which is in agreement with the results discussed above (Wang et al. 2018). Recruitment of these event representations has been found independent of sensory modalities, stressing their abstract nature (Zadbood et al. 2017). Furthermore, comprehension of different script-based stories leads to distinct sequential schematic event patterns (Baldassano et al. 2018).

Overall, there is now substantial evidence showing an overlap in brain activation for abstract and concrete concepts, not only for actions (Desai et al. 2018). We focused here on actions and how mental simulation provides a general inference mechanism across levels of abstraction, either through recruiting detailed sensorimotor representation or through relying on schematic simulation. (As a further example, Chwilla & Kolk [2005] showed a priming effect for words that are related only through a shared schematic situation, suggesting that a schematic simulation can facilitate more efficient access of information for abstract information.) In particular, we reviewed results that show, first, how embodied representations are differentially recruited in language comprehension of abstract content, and, second, how this is complemented by the recruitment of more abstract schematic action structures that are engaged in mental simulation and that support abstract inferences. This broader view of mental simulation contrasts with the authors' view of the limitations of a simulation-based account.

A modern materialist approach to abstraction, concreteness, and explanation in cognition

Richard Shillcock

Department of Psychology, The University of Edinburgh, Edinburgh EH8 9JZ, UK.
rcs@inf.ed.ac.uk https://sites.google.com/site/rcspplsinf/home

doi:10.1017/S0140525X19003066, e148

Abstract

Although endorsing the authors' concentration on the issue of abstraction, I critique (a) the philosophical nature of their abstract–concrete dimension, (b) their view of the brain–world barrier, and (c) their implicit positivist one-way hierarchy that has abstraction as the goal.

We cannot make a statement or an experimental manipulation without abstracting. Therefore, Gilead et al.'s focus on abstraction is welcome; as is their replacing the old empiricist individual – passively waiting to interpret incoming data – with the predicting subject. However, much remains of the old positivism. Explanatory theorizing in cognitive science is best served by a modern materialist approach with a fuller picture of abstraction and of *universals*.

The authors posit entities that are exclusively one kind of universal. They are all "abstract universals." For materialists, "abstract" means "sparsely connected" and "concrete" means "densely/widely connected." It is a continuum, in a monist worldview. An abstract universal (e.g., "verb," "script," and "episode") captures that which is similar between many entities. Beyond that similarity, it is relatively contentless. Reliance solely on abstract universals entails describing and redefining *ordered relations* between such entities. It necessarily leads to endless notational variants, as the authors' review shows, and eventually to each abstract universal being defeated by new data. Nevertheless, such universals provide a necessary initial traction on a domain. They may play a later theoretical role in conjunction with a second type of universal.

A *concrete universal* (the term is originally from Hegel, but cf. Vico's "imaginative universal," Goethe's "Ur-phänomen," and Vygotsky's "unit of analysis") is something *material* that speaks to and mediates everything else in the domain being studied (Shillcock 2014). Unifying other entities is the classic definition of a universal. Further, it behaves in a way that characterizes the whole domain. (Indeed, it is one approach to defining a domain.) It is the explanatory essence that provides us with the best understanding of the totality of the domain. For example, in explaining how cells come together to create the domain of bodily anatomy, the *stem cell* would be the relevant concrete universal: It is a cell like other cells but – given the necessary conditions – leads us to the totality of the domain. What might be the relevant entity in cognition, from neuron to hemisphere, from orienting reflex to spoken word? One approach to identifying it is to critique Gilead et al.'s definition of the domain.

The authors make the world–brain barrier a key division by proposing a hierarchy from the sensorium "up" to increasingly far removed and less specific mental entities. These "topmost" entities are still immaterial abstract universals; in deep learning, such topmost entities might successfully label a picture, but that is a long way from the productive agency that characterizes human cognition. There is a suspicion that a homunculus is lurking at the top.

The authors' ontology reveals the conventional positivist notion of stepwise building up, from supposedly assumptionless "atomic" foundations, verifying each move, until the highest processing is achieved. The human central nervous system certainly contains hierarchical visual processing close to the sensorium, as the authors note, but it is also characterized by enormous recurrency (with predictive processing and the incorporation of material artefacts being the most sophisticated aspects). In this sense, the authors tell a conventional story concerning the relationship between the single subject's brain and the objective outside world.

Philosophers have long claimed that people have "precipitated out" the results of cognition and that these real-world artefacts are legitimate components of cognition. We can see such "tools" in the outside world – machines, locomotives, and words. We can also talk about "tools" *inside* the brain (but not as "tools all the way down," which makes "tool" a less than useful abstract universal).

A materialist analysis develops all of these arguments to claim that the *hemisphere* is the relevant concrete universal in the domain of wider cognition (Shillcock et al. 2019). A single hemisphere is substantially capable of doing anything that a whole brain can do – its activities characterize the whole domain of the cognizing brain. Every aspect of cognition is affected by the hemispheric divide, given the extensive lateralization and specialization of function unique to the human brain. An enormously productive dialectic emerges between the "two brains in one cranium" sharing the same world, and finessing any need for a homunculus. Each hemisphere is productively predicting and modelling the other, effectively using the other as a tool. Such mutual modelling ensures a unity of conscious experience.

As a second example, the foetus/infant responds to *speech* as an "abstract" (sparsely connected) material entity. Speech constituents are internalized by a variety of actions, small and large-scale. Over time, they become more "concrete," more densely connected with each other and with different activities. Specifically for English, we might trace the life of the schwa-sound (itself a word – "a" – and a paralinguistic gesture) within each of the subdomains of language behaviour (phonology, syllabicity, syntax, semantics, and so on) as a promising concrete universal that allows speakers to negotiate between old and new information, which is the fundamental nature of spoken communication.

Selecting a candidate concrete universal in no way excludes us from researching the characteristic activities of any other entity, large or small. It might lead us to clarify the structure of a larger or smaller domain, with its own characteristic processing. A concrete universal provides us with the deepest joint at which to carve nature, to reveal the essential "logic" of the behaviour of the domain. An explanatory theory requires us to be able to move dialectically between a relatively simple material element playing the critical role in the domain – the hemisphere, the schwa-sound, in our examples – and the *totality* of the moving, acting cognitive agent, the goal being to return our theorizing to that latter totality.

Abstracting reward

David Spurrett

Department of Philosophy, University of KwaZulu-Natal, Durban 4041, South Africa.
spurrett@ukzn.ac.za https://philpeople.org/profiles/david-spurrett

doi:10.1017/S0140525X19003042, e149

Abstract

The costs of and returns from actions are varied and individually concrete dimensions, combined in heterogeneous ways. The many needs of the body also fluctuate. Making action selection efficiently track some ultimate goal, whether fitness or another utility function, itself requires representational abstraction. Therefore, predictive brains need abstract value representations.

The target article develops a theory of abstraction largely focused on representations of objects and states of the external world. It argues that predictive cognition needs an account of hierarchical representational diversity, partly to explain the human capacity to transcend the "here and now." So far, so good.

Although the topic is not explicitly addressed, the arguments concerning degrees of abstraction, and the case for a hierarchy, apply just as forcefully in the case of representations of the values of, or predicted returns from, behaviours and actions as they do to representations of the external world. (Gilead et al. say that their theory is intended to apply to "mental states" including desires and intentions, but go on to focus almost exclusively on beliefs.) Evolved cognition, after all, and whatever you think of predictive processing, isn't a goal in itself, but a means to co-ordinating behaviour with contingencies and changes in the body and world. The actions available to an agent typically have varying and multi-modal costs and returns. The costs include time, energy, depletion of specific resources such as water and salt, exposure to various risks, and the opportunity cost of forgone actions. The returns can include hydration, the many dimensions of nutrition, rest, safety, territory, and opportunities to mate. There are, that is, many concrete dimensions of cost and benefit.

At the extreme of low abstraction, specific actions can be occasioned by detections of specific concrete, single modality returns or costs (e.g., dehydration, or estimated effort). At the other, action can be selected and prioritised on the basis of wholly abstract considerations, such as utility. In evolutionary settings, utility can be taken to correspond to fitness, permitting economic and decision-theoretic analyses of rational agency to be applied to evolutionary analyses of optimality. Okasha (2013), for example, argues that an agent that acts and chooses as if performing Bayesian updating is rationally optimal, and hence a plausible theoretical target for the influence of natural selection on its dispositions and their cognitive implementation. Behavioural ecology is methodologically committed – even if only as a regulative principle – to the determinacy of the fitness return of *all* behaviour, a view expressed in McNamara and Houston's assertion that any "attempt to understand behavior in terms of the evolutionary advantage that it might confer has to find a 'common currency' for comparing the costs and benefits of various alternative courses of action" (McNamara & Houston 1986, p. 358).

Cognitively useable utility representations would need to integrate the various dimensions of cost and return, in ways sensitive to relations of substitutability between the dimensions and the ways they combine in external objects. (Individual food items, e.g., combine various nutrients, the current value of each of which will depend on the state of the body, and the varying costs of getting access to the food.) Utility representations suitable for guiding action would need to do the same for the various capacities of the body itself, because actions are more or less substitutable with each other, and some goals achievable by more or less large collections of different deployments of the powers of the body. Arguably, some efficiencies in allocation can only be achieved by relatively abstract value representations (Spurrett 2019).

It isn't clear whether our utility representations are fully abstract. Optimists on this question, especially some neuroeconomists, think there's good evidence that humans and some non-human animals process highly abstract utility representations for rewards in widely varying modalities. Levy and Glimcher (2012), for example, survey studies finding consistent neural signatures of the value of monetary gains and losses, cumulative monetary rewards, anticipation of varying monetary rewards, expected values of uncertain monetary rewards, and discounted value of delayed monetary rewards. More tellingly, they review studies involving choices with at least one incentive other than money, including consumer goods, gustatory rewards (water, juice, and food), physical pain, and social reputation, still finding the same general neural signature. Those who are less optimistic point to the ways in which learning about rewards exhibits failures of abstraction. Rolls (2013), for example, argues that Levy and Glimcher's evidence doesn't decide between a situation where there is a genuine "common currency" (fully abstract utility) and one where different rewards compete on a "common scale" whereas retaining important differences in their links to specific rewards and courses of action. We've long known, furthermore, that the form of conditioned behaviours is related to the reward they deliver, for example that pigeons peck a key leading to water delivery with a drinking-appropriate action, whereas a key leading to food elicits an eating peck (Jenkins & Moore 1973).

We don't need to settle the question of whether human value representations are fully abstract here. What matters is that abstraction applies to costs and returns, and our value representations are abstract to a fairly high degree. This is, furthermore, crucial for useful mental time travel (transcending the "here and now"). If time-travel is to pay its way, as it must have for selective processes to have designed it, it must have contributed to the general function of cognition. That is to say, agents transcend the here and now in order better to determine what to do, here and now. Prospection without reasonably accurate estimation of the costs and returns from the implied courses of action, including extended chains of them, is frivolous speculation.

The neglect of these issues in a target article with "predictive brain" in the title is striking partly because it was in neuroeconomics that the first compelling evidence of prediction-error-based processing in natural brains was found. Even if, in the theoretical limit, reference to reward is replaced by talk of "hyper priors" or some other term from the predictive processing framework, it's still important to recognise and take account of the ways in which abstraction is significant for reward based or evaluative representations, and how such representations are important for overall organismic efficiency.

A challenge for predictive coding: Representational or experiential diversity?

Martina G. Vilas[a] and Lucia Melloni[a,b]

[a]Department of Neuroscience, Max Planck Institute for Empirical Aesthetics, 60322 Frankfurt am Main, Germany and [b]Department of Neurology, NYU Comprehensive Epilepsy Center, School of Medicine, New York University, New York, NY 10016.
martina.vilas@ae.mpg.de martinagvilas.github.io
lucia.melloni@ae.mpg.de
aesthetics.mpg.de/en/the-institute/people/lucia-melloni-en.html

doi:10.1017/S0140525X19003157, e150

Abstract

To become a unifying theory of brain function, predictive processing (PP) must accommodate its rich representational diversity. Gilead et al. claim such diversity requires a multi-process theory, and thus is out of reach for PP, which postulates a universal canonical computation. We contend this argument and instead propose that PP fails to account for the experiential level of representations.

Gilead et al. propose that the brain embodies a hierarchy of abstraction processes whose levels differ not only in *what* they are representing (its inputs), but also in *how* they impose constraints in their interactions with other representations. In other words, they propose a structured system of symbolic mental representations that allow for various distinct algorithmic operations. They further argue that because predictive processing (PP) neglects the importance of symbolic processing and its functional heterogeneity, it cannot account for the mind's representational diversity and combinatorial nature.

We agree with the authors that any all-encompassing theory of brain function needs to accommodate the brain's representational diversity. We argue, however, that in hierarchical PP this is captured by deep and less contextualized generative models encoded in the neural system, whose acquisition and update are based on a unifying inferential principle (Friston et al. 2017a; Melloni et al 2019; Snyder et al. 2015). The generative model probabilistically represents our beliefs about the hidden states that give rise to our sensory experiences, in order to minimize surprise (Friston et al. 2017a). Central to our point, the model can be ascribed key characteristics that make it suitable to explain the mind's representational diversity (Melloni et al. 2019).

First, the generative model can represent the probabilities of both discrete and continuous events (Friston et al. 2017a). In the former case, where events in the world can be categorical (e.g., we can be either dead or alive but not both at the same time), the corresponding beliefs would be represented as a probability distribution over a finite set of states. In contrast, continuous events (e.g., a moving object at a particular place in the visual field) would be encoded as an analogous probability density. The kinds of generative models suited for the higher-level abstract representations introduced in the article, such as categories, would thus rely on discrete states of the world that can be described with symbolic or semantic labels.

The second relevant feature is that the generative model can be expressed as a hierarchical Bayesian graph with nodes and edges, where the nodes represent the hidden states and the connections stand for the conditional dependencies between them (Friston et al. 2017a). In this context, the network structure of the hierarchy of abstractions postulated by Gilead et al. can be accounted for.

Third, the generative model can be a *deep temporal* nested hierarchy (Friston et al. 2018) allowing mental simulation of possible future states during decision making (i.e., mental time travel). This third property also gives the neural system the capability of making predictions at multiple scales of abstraction.

Finally, the generative model can take a factorial form, in which diverse causes are represented as independent and separate states that can be brought together (e.g., through convergent connectivity) to explain the sensory input at hand and/or produce a new representational outcome (Friston & Buzsáki 2016). An advantage of factorizing the generative model is to reduce combinatorial complexity, as the system does not need to explicitly code for all possible combinations of states. For instance, a visual event can contain any arbitrary combination of objects (a "what" attribute), their location ("where"), and a timestamp ("when"). The brain could factorize those attributes such that they can be put together to represent every possible event (Auksztulewicz et al. 2018; Friston & Buzsáki 2016). A similar scheme can be devised for the case of higher-order representations, in which predicators and other abstracta could be separately represented in the graph, and the result of their combination can be obtained *via* convergence of the associated nodes.

All in all, if we take the described properties into account, it becomes evident that the generative model can place structural constraints that impact the kind of neuronal computations allowed by the representational system (Melloni et al. 2019).

How this structure is acquired and implemented in the brain, however, remains a question to be addressed (the structure learning problem, see Griffiths et al. 2010; Melloni et al. 2019). Regardless, it can still be argued that PP, at the computational level, can accommodate symbolic representations: these are embedded in the generative model.

But, is the account of the representational diversity enough for PP to become a unifying theory of the brain? We argue that this is not the case, as PP still neglects the mind's diversity at the *experiential* level. The theory fails to explain why and how the processing of different representations "feels" the way they do. In our daily life, clear-cut qualia discontinuities appear between modality-specific representations that differ in their sensory input. For example, "seeing" a dog does not evoke the same subjective experience than "hearing" it barking. How this experiential distinction can be explained in terms of neural processing is ignored by current PP theories. Along the same lines, PP fails to differentiate between conscious and unconscious events. A hierarchy of abstract representations is also evident in the case of unconscious predictions carried out automatically during perception, like pattern completion processes (Schwiedrzik & Freiwald 2017). Why those automatic predictions do not reach consciousness is not yet articulated by PP. To state the problem in broader terms, why certain processes are accompanied by consciousness and others are not, how differences in qualia translate to differences at the neural processing level, and how subjective experience arises from neural signals, is currently not spelled out by PP models. Of note, diversity at the experiential level cannot be tackled by mapping different qualia to distinct cortical areas or brain networks. A map by itself does not explain why or how experiential diversity comes about (Poeppel 2012).

To conclude, although we agree with the authors that the mind's representational diversity imposes challenges to PP, we propose that it is at the level of the experience that PP fails, and not at the level of the functional diversity, which is within reach of the hierarchical and deep temporal generative models.

The role of sleep in the formation and updating of abstract mental representations

Antonino Visalli[a] and Nicola Cellini[a,b,c,d]

[a]Department of General Psychology, University of Padova, 35131 Padova, Italy;
[b]Department of Biomedical Sciences, University of Padova, 35131 Padova, Italy;
[c]Padova Neuroscience Center, University of Padova, 35121 Padova, Italy and
[d]Human Inspired Technology Center, University of Padova, 35131 Padova, Italy
antonino.visalli@unipd.it
nicola.cellini@unipd.it

doi:10.1017/S0140525X19003005, e151

Abstract

According to Gilead and colleagues, to be efficient abstraction requires a hierarchical organization of information into long-term memory. But, how and when are abstract representations consolidated into long-term memory and how are they integrated with pre-existing abstracta are questions not discussed by Gilead and colleagues. Here, we propose that these processes occur preferentially during offline periods such as sleep.

We agree with Gilead et al. on the necessity of a hierarchical organization of abstract representations in long-term memory (LTM) for the efficient construction of new mental representations. Although this aspect is merely mentioned in the target article, it is closely related to a core assumption of predictive brain theories. Specifically, as a predictive machine, the brain should encode new information by "explaining away" discrepancies between incoming signals and top-down predictions based on prior knowledge (internal models) acquired in the past (Clark 2013). Then, discrepancies can be used as bottom-up error signals to update internal models. It follows that a deeper integration between Gilead and colleagues' conceptualization and the predictive brain framework requires to specify abstractions in terms of updated representations of previously stored abstractions. Moreover, new abstractions need then to be consolidated and integrated into the complex and hierarchically structured representations stored in LTM to be used for future acts of abstraction.

Based on these theoretical considerations, we want to highlight the importance of making a distinction between *online* and *offline* abstraction processes. Online abstractions can be seen as short-term inferential processes that, similarly to the ones mainly described in the target article, lead to new (updated) mental representations by integrating evidence – whose content is driven by ongoing internal/external experience – with prior abstractions retrieved from LTM. The short-term status of online abstractions is fundamental to avoid the simple "hoarding" of representations in LTM, which would lead to a not efficient organization of internal models. Conversely, efficient integration of new representations with pre-existing *abstracta* requires a deep reorganization of pre-existing representations to maintain coherence in such internal models and to ensure their predictive power by preventing redundancy and overfitting (Pezzulo et al. 2017). Therefore, some key questions not discussed by Gilead and colleagues need to be addressed: How and when are abstract representations consolidated into long-term memory? And how are they integrated with pre-existing abstracta? Here, we posit that the complexity of such reorganization requires offline acts of abstraction during which the brain – decoupled from ongoing experience and in the absence of new evidence – recursively generates and evaluates inferential updates to improve the hierarchical organization of the internal models (Pezzulo et al. 2017). Grounding on compelling evidence that has emerged in the past two decades about the active role of sleep in (i) consolidating spatial, episodic, and semantic memories (Rasch & Born 2013), (ii) facilitating the extraction of statistical regularities, gist, and overarching rules (Stickgold & Walker 2013), and (iii) supporting the formation of internal predictive models (Lutz et al. 2018), sleep appears to be the ideal status where offline abstraction processes may take place.

More intriguingly, we propose that the sequential interplay between non-rapid eye movement (NREM) and REM sleep, which iteratively interleaves in 90-min cycles during nocturnal sleep, may be critical for the development of the representational hierarchy proposed by Gilead and colleagues by promoting both the consolidation and the updating (reorganization) of these internal models (Ballesio & Cellini 2019). Indeed, NREM and REM sleep show very different patterns of brain activity, which likely support complementary functions in structuring, consolidating, updating, and minimizing the complexity of internal models (Hobson & Friston 2012). In particular, NREM sleep may promote an initial consolidation of newly encoded information into pre-existing memories that are coherent with, and supportive of, our prior beliefs and knowledge (Oudiette & Paller 2013). Compelling evidence indicates that in NREM2 and NREM3 the hippocampus controls neural replay in the cortex ensuring that memories related to a prior representation are replayed concurrently (Klinzing et al. 2019). This overlapping replay seems to lead to the abstraction of shared properties although redundant information is pruned (Feld & Born 2017; Lewis et al. 2018; Stickgold & Walker 2013). Once coded in the neocortex during NREM sleep, such representations might serve then, as *concreta* for further abstraction processes during REM sleep. Indeed, during REM a reduced hippocampal input to neocortex is coupled with a massive neural activity (driven by ponto-geniculo-occipital waves) that can randomly activate cortical representations. This highly active state might allow scanning memories to find meaningful regularities between events, and to generate and strengthen new abstractions between both cortical representations activated during the NREM and other, apparently not related, representations (Lewis et al. 2018; Llewellyn 2016). Moreover, it has been proposed that during REM the sleeping brain tends to simulate and rehearse fictive scenarios to produce prediction errors and, therefore, refine (i.e., update) the internal models before the next wakefulness (Friston et al. 2017b; Hobson et al. 2014).

Finally, we believe the different nature of neural activity patterns and associated cognitive processes observed during NREM and REM sleep may support the qualitative distinction between *regular* abstracta and predicators. According to Gilead and colleagues, regular abstracta can be seen as saturated representations that arise from regularity and/or shared properties between concreta. Conversely, predicators are unsaturated entities that behave like functions to create new associations between representations. This distinction emerges also in the abstraction processes described above. Indeed, during NREM abstractions emerge from the overlap of shared properties between representations, whereas in REM new associations between representations are instantiated. Therefore, beyond the qualitative difference between abstract representations, we propose that future studies may take advantage of investigating qualitative differences in abstraction processes such as those promoted by the different physiology and neural mechanisms that the brain experiences in NREM and REM sleep.

In sum, we believe that differentiating between online and offline (particularly sleeping) abstraction processes is fundamental for understanding how the representational hierarchy proposed in the target article is built up and optimized for prediction. Moreover, based on the different neural activities observed in NREM and REM sleep, we suggest that the identification of qualitative different abstraction processes (beyond the qualitative difference in representations) will offer a key to unveil the representational bases of cognition.

Acknowledgments. This work was supported by the University of Padova under the STARS Grants program to N.C. The present work was carried out within the scope of the research program "Dipartimenti di Eccellenza," which is supported by a grant from MIUR to the Department of General Psychology, University of Padua.

Successful simulation requires bridging levels of abstraction

Zidong Zhao[a], Judith N. Mildner[a] and Diana I. Tamir[a,b]

[a]Department of Psychology, Princeton University, Princeton, NJ 08540 and [b]Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08540.
zidong@princeton.edu
jmildner@princeton.edu
dtamir@princeton.edu

doi:10.1017/S0140525X19003133, e152

Abstract

Although many simulations draw upon only one level of abstraction, the process for generating rich simulations requires a dynamic interplay between abstract and concrete knowledge. A complete model of simulation must account for a mind and brain that can bridge the perceptual with the conceptual, the episodic with the semantic, and the concrete with the abstract.

Gilead and colleagues operationalize abstraction as a spectrum from the concrete and modality based, to the intermediate and multimodal, to the most abstract and theory-like. This aligns with earlier work on the spectrum of representations that people draw upon, from episodic to semantic (Tulving 1972), or from model-free to model-based (Gershman & Daw 2017). Gilead et al. further suggest that higher-level abstractions are representationally and neurally distinct from lower-level ones, and that different types of mental travels rely on different levels of this abstraction spectrum. Specifically, Gilead et al. limit simulation to more concrete representations. This formulation successfully captures some cases of prospection. For example, it accurately predicts that people will rely on abstract and theory-based sources in order to mentally travel to distal events (e.g., events in the far future, or about dissimilar others) and more concrete sources for proximal events (e.g., events in the near future, or about similar and familiar others). However, this formulation, like much prior work on simulation, underspecifies the dynamic ways in which different levels of abstraction interact during simulation. We propose that simulation requires agents to bridge representations at multiple levels of abstraction, in a dynamic interplay between abstract and concrete knowledge.

Higher-level abstracta scaffold lower-level ones, across multiple domains. During visual perception, category cues shape perceptual expectations (Gandolfo & Downing 2019). During social perception, the structure of emotion concepts shapes the perception of emotion in faces (Brooks & Freeman 2018; Brooks et al. 2019). During simulation, abstracta such as event knowledge and personal semantics (Renoult et al. 2012) shape which specific, concrete details people will use to fill in the blanks (Conway & Pleydell-Pearce 2000; D'Argembeau & Mathy 2011). This suggests that abstract and semantic knowledge support episodic simulation.

To test this perspective, our lab studied individuals who excel at simulation: creative experts. Most people produce distal simulations with highly abstract, schematic content, as predicted by

Gilead et al. However, creative experts continue to generate rich, vivid, detailed simulations even about very distal events. How do they do so? We find that during distal simulations, creative experts preferentially recruit the dorsomedial subsystem of the default mode network (Meyer et al. 2019), a network associated with semantic knowledge and high-level construals (Baetens et al. 2014; Binder et al. 2009; Fairhall & Caramazza 2013; Gilead et al. 2014; Simony et al. 2016; Spunt et al. 2016); both groups recruit the medial temporal subsystem, associated with episodic detail, to the same extent. This suggests that creative experts are able to generate vivid and concrete distal events by harnessing abstract knowledge. In fact, the system that organizes abstract information may facilitate creative ability: semantic memory organization in creative experts is less hierarchical than in typical thinkers (Kenett et al. 2014; Mednick 1962). Together, these findings suggest that successful simulation requires the ability to flexibly integrate abstract knowledge with concrete representations.

Not only do abstract representations shape concrete ones, concrete representations also shape abstract ones. Abstracta are not stable knowledge stores; they are dynamic, probabilistic distributions over concreta (Griffiths et al. 2012). Abstracta shift as the availability of concreta ebbs and flows. This has been demonstrated across different domains of cognition. During decision making, an incidental reminder of a single past reward episode biases decisions away from abstract summary reward values (Bornstein et al. 2017). During impression formation, people's general impression of others' moral character rapidly updates after observing a single new moral act (Siegel et al. 2018). During trait judgments, people retrieve behavioral exemplars in order to estimate new traits in the absence of pre-existing trait representations (Klein et al. 1992; Markus 1977). During spontaneous thought, concrete perceptual thoughts cue subsequent abstract thoughts (Bar et al. 2007; Klinger 2013; Mildner & Tamir 2019; Northoff 2018). Finally, during event simulation, heightened access to specific, concrete details shifts people's evaluation of abstract event features such as valence (Jing et al. 2017; Madore et al. 2019). Thus, concrete representations alter existing abstract representations, and can even be used to generate new ones on the fly.

When the ability to successfully bridge abstract knowledge with concrete knowledge is impaired, distal simulation is unsuccessful. For example, patients with depression show deficits in retrieving concrete, specific details of positive episodic memories (Williams & Scott 1988). These patients also show deficits in updating negative abstract representations from new, concrete instances to the contrary (Korn et al. 2014), and in producing vivid prospections about future positive events (Gamble et al. 2019). These patients have access to abstract knowledge, but without concrete details, they are stymied in their attempts to simulate successfully. In contrast, patients with deficits in semantic memory have access to concrete episodic details (Irish & Piguet 2013). Yet, these patients likewise struggle to simulate novel events. Instead, they tend to simply recall past events when asked to imagine a future event (Irish et al. 2012). These patients have access to concrete details, but, without abstract knowledge, they fail to recombine these details into novel simulations (Irish & Piguet 2013). Together, these findings suggest that successful prospection relies on the flexible interplay between abstract and concrete representations.

Although some simulations draw upon only one level of abstraction, as Gilead et al. propose, rich simulations often require a dynamic interplay between multiple levels of abstraction. Abstract representations guide the search for detailed concrete representations, and concrete representations are used to construct abstract representations. From this perspective, we interpret the overlapping brain regions associated with semantic cognition and prospection as a manifestation of such reciprocal, dynamic processes. We hope future research will build toward a more complete model of simulation, one that can accommodate a mind that dynamically bridges the perceptual with the conceptual, the episodic with the semantic, and the concrete with the abstract.

Authors' Response

Above and beyond "Above and beyond the concrete"

Michael Gilead[a], Yaacov Trope[b] and Nira Liberman[c]

[a]Department of Psychology, Ben-Gurion University, Beersheba 84105, Israel; [b]Department of Psychology, New York University, New York, NY 10003 and [c]Department of Psychology, Tel-Aviv University, Tel-Aviv 69978, Israel.
mgilead@bgu.ac.il http://www.gileadlab.net/
yaacov.trope@nyu.edu http://www.psych.nyu.edu/tropelab/
niralib@tauex.ac.il https://en-social-sciences.tau.ac.il/profile/niralib

doi:10.1017/S0140525X20000023, e153

Abstract

The commentaries address our view of abstraction, our ontology of abstract entities, and our account of predictive cognition as relying on relatively concrete simulation or relatively abstract theory-based inference. These responses revisit classic questions concerning mental representation and abstraction in the context of current models of predictive cognition. The counter arguments to our article echo: constructivist theories of knowledge, "neat" approaches in artificial intelligence and decision theory, neo-empiricist models of concepts, and externalist views of cognition. We offer several empirical predictions that address points of contention and that highlight the generative potential of our model.

We were fortunate to receive numerous enlightening commentaries from a broad spectrum of researchers in social, cognitive, and developmental psychology, philosophy, neuroscience, and artificial intelligence. The commentaries revisit some of the classic debates concerning mental representation and abstraction and inform current models of predictive cognition. Below, we respond to the central issues raised by the commentators.

R1. "Neats" and "Scruffies"

Several commentators (e.g., **Fields & Glazebrook**; **Davis, Altmann, & Yee** (**Davis et al.**)) oppose the necessity for the distinctions we make between (supposedly) different representational entities. Such commentaries echo the classic debate in cognitive science between "Neats" and "Scruffies" (see Schank & Abelson, in Clark 2013; Marcus 2009).

For example, **Fields & Glazebrook** believe that it would be possible to find "a single computational architecture (that) works… at every useful level of analysis." Such models would have "the advantage of being rigorously testable at every experimentally-accessible level of analysis, from those of basic physics, intracellular, and cellular processes up to the whole-organism scale and beyond." As examples for such "master algorithms," they mention Karl Friston's Active Inference Theory and their own "Chu space" model.

Undoubtedly, the idea of a "theory of everything" is attractive, if tenable. However, the suggestion that the same science can apply to every level of analysis – from physics to cellular processes, to whole organisms and societies – is an extraordinary claim that would require extraordinary evidence. If a universal model can be used to predict hurricanes and anticipate wars, then it could be argued that specific models in different sciences may become redundant. In the meantime, each of the sciences still requires its own models (e.g., political science models for prediction of war, biological models for prediction of cancer, and so on). All in all, we believe that sciences have progressed, and have done so without a theory of everything.

Furthermore, we believe that general and specific theories need to complement each other to provide humans with an understanding of their world. A good analogy is the Darwinian theory of evolution that provides a "neat" overarching framework for the understanding of life, and fully accommodates the "scruffier" world of genetics. It is the synthesis between these two lines of research that provided a real understanding of biological life. It appears that our position does not differ from that of **Friston**, whose Active Inference Theory is cited by **Fields & Glazebook** as the prime example of scale-free models. As Friston notes:

> The generative model is important because most of the heavy lifting – in terms of understanding structure–function relationships in the brain – rests on its form. In other words, if one knows the generative model, model inversion can be cast in terms of the Bayesian brain hypothesis (in a normative sense) (Doya 2007; Knill & Pouget 2004) or combined with standard inversion schemes to generate neuronal processes theories about computational brain architectures and neuronal message passing (Friston et al. 2017c).

In other words, although broad "scale-free" theories such as Active Inference Theory (and like evolutionary theory) may describe universal principles of biological entities, they are still "unsaturated," and their application to a given domain requires filling in many "scruffy" details. A full understanding of cognition will entail having an accurate picture of different generative models (i.e., representational entities) used by the brain, as well as its overarching organizing principles.

Importantly, understanding the universal principles that govern a system does not tell us all there is to know about the system. The system will be influenced by random, as well as systematic factors, and by forces that are external to it. For example, the theory of evolution does not entail that there should be lions and giraffes in the world. Nonetheless, these specific implementations of the evolutionary algorithm (e.g., the existence of giraffes) are indispensable to our understanding of that which exists in our world.

R2. Filling in the blanks

Ironically, our criticism that overly-broad models are insufficient to account for the richness of cognition has been directed against our own account.

Fiedler suggests that our model does not sufficiently discuss how behavior is determined by extra-cognitive factors, namely, how the structure of the world shapes psychological phenomena. For example, if in one's experience, people who are similar to us occupy more spatially proximal places than people who are different than us, one should see various effects on human behavior that reflect this association between social distance and spatial distance. An explanation of this association may not require a lengthy discussion of humans' motivations and neural mechanisms. Instead, this phenomenon may be well-explained by the analysis of statistical regularities of the environment to which they were exposed.

We agree that distance-to-distance associations can be found in one's environment. We think, however, that the association between distance and abstraction may have a different status. The variability and uncertainty that come with increased distance necessitate abstract mental representations in an individual's mind only if she seeks to travel the distance whereas still predicting events in the new environment. To use **Fiedler**'s example, the fact that a map that represents a larger area (e.g., a continent rather than a neighborhood) would have cruder resolution is dictated by constraints on the usefulness of this map for a traveler, rather than being a physical necessity. Therefore, we agree that objects at a greater distance entail greater uncertainty, and we think that this regularity gives rise to a stable *intrapsychic* link between construal and distance.

Indeed, our account takes an intrapsychic perspective of human psychology. However, this does not mean that we (or the predictive cognition view more broadly) ignore the structure of the world. Instead, it means that biologically inspired accounts of human psychology choose to focus on ecologies within which we were embedded for the longer timescales. The long neck of the Giraffe is an imprint of its environment over evolutionary timescales. Similarly, the representational structures we discuss are a reflection of the ecology that has impinged upon our mind and brain.

This point is also highlighted in the commentary by **Friston**, which discusses how the "factorization" of the mind into different representational types is a consequence of the structure of the world. As Friston writes: "knowing what something is, does not tell you where it is and vice versa. This (conditional) independence is manifest beautifully, in terms of the functional anatomy of the dorsal and ventral streams in the brain."

Indeed, the distinction between modality-specific, multimodal, and categorical abstracta can be thought of in terms of the *world* rather than the *mind*.

Modality-specific visual representations are an adaptation to the fact that objects in the world reflect light at various frequencies. An organism living in a world of complete darkness for several millennia may lose the functionality of V1 and eventually may even lose the need for eyes altogether. Multi-modal representations are a function of the statistical regularities embedded in the world. The location of objects in space and time is not random, because things that are closer in time are also closer in space. An organism living in a world wherein light shines for only a few seconds every several minutes will not develop sophisticated mechanisms for binding visual sensations. Finally, a system for categorical representation probably would not have evolved if proto-humans did not have an anatomy that allowed them talk to each other.

Surely, aside from the stable characterizations of our ecology, there are aspects of the world that influence the way we represent it – without entailing any changes in neuroanatomy. When a psychologist is confronted with a specific behavior, it is undoubtedly

https://doi.org/10.1017/S0140525X19002000

essential to consider the degree to which this behavior is well-explained by factors that are external to humans' internal makeup.

Similarly to **Fiedler**, **Demetriou** suggests that our model misses out on some crucial details. He sees deficiencies in that that we do not provide: (i) a mechanistic description of the process of abstraction, (ii) a detailed developmental theory of the emergence of different mental travel capacities, and (iii) an account of individual differences in abstraction and mental travel.

We agree with **Demetriou** that the model we describe leaves these crucial questions open. However, we see this as a question of scope, rather than an inherent limitation. We think that our framework has the potential to consolidate further ideas and findings in different areas of investigation, such as those highlighted by Demetriou. For instance, although our account does not provide a detailed mechanistic model of the act of abstraction, it provides a framework for such details to be filled in. As we suggest in the manuscript, the broad principles we outline take different forms in different modalities.

Scharinger's comment fills some of these missing specifications by explaining how the hierarchical nature of representations is expressed in the domain of sound and speech. His analysis further highlights how a hierarchy of abstractness is an omnipresent regularity of cognition rather than a feature of a specific system (e.g., the visual system).

Furthermore, as noted by **Demetriou**, our account does not provide a detailed developmental trajectory of the emergence of different aspects of mental travel. The commentary by **O'Brien, Rutt, & Atance** begins to fill in these details by describing current evidence about children's ability to traverse different dimensions of psychological distance and highlights future efforts to connect these data with children's emerging representational capacities. As they note: "The results of such efforts may not fundamentally contradict Gilead et al.'s claims, but may instead add richness and precision to an account of how humans develop the remarkable capacity for mental travel."

Finally, as noted by **Demetriou**, our account does not provide a discussion of individual differences. The commentary by **Belmonte** highlights how the model is consistent with observations concerning variability in autistic traits, and how the relevance of our model to this literature can be empirically tested.

Belmonte's compelling theory of autism coheres with our view of representational hierarchy and its role in prediction. This account explains how many of the cognitive impairments observed in autistic individuals (e.g., inability to understand the minds of others, to consider temporally distant outcomes of one's behaviors, and to deal with uncertainty) can be thought of as deficiencies in mental travel, which in turn, can result from autistic individuals' over-reliance on modality-specific representations, rather than multi-modal and categorical abstracta (which is evident, e.g., in autistic individuals' preference for processing modality-specific relationships, described in the research on the "systemizing" phenotype). Belmonte's account addresses the emotional burden of being unable to predict the world and the futility of trying to find structure in a world of constant change using low-level, concrete mental representations.

Belmonte also suggests new predictions of this account linking abstraction, prediction, and autism. One interesting ramification of this link between our model and Belmonte's may relate to a better understanding of the diversity of autism. It is often argued that autism is, in fact, a collection of disorders; however, the underlying endophenotypes that give rise to different manifestations of autism are difficult to identify without a generative theory of this disorder. If it is indeed the case that autism is related to over-reliance on concrete representation, it is possible to ask, for example, whether some autistic individuals overly rely on multi-modal abstracta (but have difficulties with forming and using categorical abstracta) and whether other autistic individuals cannot even rely on multi-modal abstracta in processing the world. Such research may be able to relate behavioral phenotyping to the neuroscientific phenotyping of modality-specific areas of the brain, the default network, and frontotemporal regions supposedly involved in categorical abstracta.

R3. Can Predictive Processing accommodate abstract cognition?

Our paper explores the potential role of abstract (including symbolic) mental representations in predictive cognition. As we discuss in the article, one of the most prominent perspectives in the predictive brain literature is the so-called Predictive Processing (PP) approach. Several authors have responded to the target article by assessing the extent to which the representational ontology we describe is (or can be) consistent with PP.

Litwin & Milkowski present a critical perspective on the PP approach to psychiatric illness. They suggest that our account – that highlights the importance of structured representations – is at odds with the connectionism-inspired PP paradigm.

The question of whether and how our approach can be made compatible with the PP approach is somewhat tricky because there is diversity in PP models. In many cases, the exact theoretical commitments of the different models (e.g., active inference/predictive coding/the Bayesian brain) diverge. Active Inference Theory, which is the predominant account in the PP world, is a broad framework that accommodates many concrete "process models" of cognitive architecture and neuroanatomical implementation. Some of these process models deny the existence of structural representations of the kind we assume, and as such, are radically inconsistent with our view. Others are probably entirely consistent with our model.

Surely, as noted by **Litwin & Milkowski**, many PP process models of psychiatric illness describe a world that does not provide a place for the representations we posit in our ontology. This argument is actually the main point we tried to make in our discussion of the relevance of our model to PP theories. We echoed the standard psychological view of psychopathology, which assumes that abstract, structured representations of the self and others have a critical role in mental illness. Furthermore, the standard account posits that altering these representations via language-mediated processes reduces distress. Such high-level processes are not part of PP process models of psychiatric illness. Nonetheless, we suggested that abstract cognition must play a role in disease, and if that is the case, then something is missing from the PP literature on the topic.

PP models of psychiatric illness can, in principle, integrate abstract cognition into their "factorized" system of generative models. Indeed, as noted in the response by **Friston**, whereas the simplest PP process models may not be able to account for the brain's representational diversity, more sophisticated schemes of variational message passing may be fully congruent with a symbolic representational view. In our discussion of the Robot Herbie, we tried to appeal to such a formulation wherein PP models and symbolic architectures can co-exist.

https://doi.org/10.1017/S0140525X19002000

More broadly, unlike **Litwin & Milkowski**, we do think that our model coheres with (what we see as) the fundamental tenets of Active Inference Theory. Specifically, our account agrees with PP models regarding the crucial role of generative models/representations, the componential and hierarchical nature of these models (i.e., their "factorization"), and the idea that these models serve one ultimate purpose: to anticipate events – from perceptions to complex social behaviors – before they occur.

The potential ability of PP-inspired process models to accommodate higher cognition is also addressed in the response of **Vilas & Melloni**. However, Vilas and Melloni's account, similarly to **Litwin & Milkowski**'s response, presents a critical analysis of PP, and specifically, criticizes the ability of this framework to explain the phenomenon of consciousness. Villas and Melloni argue that aside from its representational diversity, the mind also exhibits phenomenological diversity (e.g., the sense of consciousness differs from that of unconsciousness); they suggest that this diversity is not captured by explicating different types of mental representations, or by describing the process of active inference.

However, the response of **Friston** to our target article does address the emergence of phenomenology, as the consequence of active inference processes upon complex representations. As he notes:

> The premise of Gilead et al. rests upon integrating influential theories in the "predictive brain camp" with "prospection (or future oriented mental time travel)." This is a big move because it entails generative models of dynamics, narratives or trajectories – with representations of the past and future. In turn, this enables the representation of states that have not yet been realised. These states undergird "simulation of future events" (intro., para. 4) and a sense of agency. In other words, the notion of a model that can "generate a representation that models the specific problem at hand" (sect. 3.1, para. 3) is exactly a generative model of the future, with "my" action as a latent state that has to be inferred.

In other words, according to **Friston**'s theory, the sense of *agency* and the act of *mental travel* are intertwined; moreover, this account implies that the *extent* of a sense of agency may relate to the extent of mental travel. Namely, asking "what is likely to be out there" is associated with a lower sense of agency than asking "what is likely out there given what is likely that I will do," and "what is likely to be out there given what is likely that I will do, given what is likely to be out there …" and so on.

This conceptualization may address a crucial requirement for a scientific theory of consciousness, namely, an explanation of the mechanisms that should give rise to the phenomenology of being an agent in the world (or not). According to **Friston**, the demarcation of a sense of consciousness may be treated as a continuous variable that pertains to the "temporal thickness" of the generated model. Moreover, the neural correlates of consciousness may be evaluated once some measure of the temporal extent of acts of mental travel is considered. This reasoning suggests that studying the process of mental travel will be crucial for studying the biological basis of consciousness. As noted, we argue that understanding the biological basis of mental travel relies on a clear understanding of different representational bases that subserve it.

Undoubtedly, much further research is needed to identify whether the principles of PP indeed extend to the domain of higher-order cognition and conceptual thought. **Petro & Muckli** describe possible routes by which to achieve this goal. Specifically, they highlight how high-resolution laminar fMRI, used in conjunction with machine-learning methods, can be used to go beyond the coarse anatomical characterization we describe, and provide a higher-resolution analysis of predictive processes at the cortical levels. They suggest focusing on the process of mental imagery as a potential central paradigm for studying conceptual processes and reconciling them with what is already known regarding perceptual processes of predictive cognition. The methodological advances highlighted in Petro and Muckli's response provide an exciting avenue for future research.

Another example of the potential usefulness of the integration we proposed relates to the commentary of **Deręgowski & Tatler**. As noted in the manuscript, predictive cognition approaches suppose that before we engage with the world, we already have a representation of how we expect it to be. Bayesian formulations of PP posit that these prior representations are empirical priors, namely, grounded on what we have previously learned. However, Deręgowski & Tatler ask how can such a claim be reconciled with research on perceptual illusions. For example, people who move around in the Ames room illusion seem to be changing their size. The prediction that people can spontaneously shrink does not seem to reflect our empirical priors, which appears to be a problem for predictive cognition theories of perception.

We think that our model of abstraction and prediction can help explain the supposed contradiction raised by Deręgowski and Tatler. Our account highlights the plurality of mental representations/models that are used by predictive cognition; moreover, it is a constructivist account, in the sense that it highlights that prediction (or more broadly, mental travel) entails forming a specific target representation from a multitude of source representations – which is an act of abstraction in and of itself. Abstraction, as we conceive it, is a judgment about which dimensions are relevant (and which are not) in a specific context.

When a constructivist perspective is applied to the Ames room illusion, it becomes more apparent that there are likely multiple dimensions and possible interpretations in the visual scene. Indeed, the hypothesis that individuals can shrink and grow is an unlikely one, but similarly, the hypothesis that the room is a trapezoid (as is the case in this illusion) is likewise unsupported by past experiences. Although *humans* don't shrink, *things* can shrink (after all, some objects, such as balloons and *Antman* do indeed shrink). In other words, it may be the case that when two unlikely hypotheses are entertained, individuals try to find meaning by generating a more abstract construal (i.e., "things shrink") that can accommodate the data. Although people haven't encountered shrinking people and trapezoid rooms, they may have encountered more shrinking *things* than trapezoid things.

R4. Are multi-modal abstracta indeed distinct from modality-specific abstracta?

A common criticism leveled against our model pertains to our claim that multi-modal abstracta and categories are functionally and neuroanatomically distinct from modality-specific abstracta. Namely, criticism that comes from proponents of embodied cognition. This view was echoed, at least to some extent, in the response of **Schilling, Chang, Rohlfing, & Spranger** (**Schilling et al.**); **Borghi & Tummolini**; **Davis et al.**; and **Pazzaglia & Leemhuis**.

The relevant question here is that: Are there mental representations that do not *necessarily* involve modality-specific properties?

https://doi.org/10.1017/S0140525X19002000

We argued that the answer to this question is a clear yes. We do not negate the existence of modality-specific representations and their importance – we only claim that there *are* mental representations that are abstracted away from their modality-specific origins. As such, an affirmative answer to the question above is quite easy to defend.

This conclusion can be justified based on conceptual analyses that date back to Kant, but psychologists such as us are better equipped to illustrate the empirical case.

As we note in the manuscript, there is ample evidence that: (i) a large part of the brain (i.e., the default mode network, including the left inferior frontal gyrus, angular gyrus, lateral temporal cortex, and anterior temporal lobe) plays a critical role in conceptual processing; (ii) these regions treat modality-specific representations of a given concept as substitutable (e.g., they respond similarly to the concept "dog" regardless of whether the participant is presented with the sound of a bark or the image of a dog's tail). Namely, they do not subserve a specific modality such as sight or sound.

To the best of our knowledge, these empirical findings are not disputed by those who endorse the embodied cognition view. A potential point of contention is whether these regions should be considered as subserving *multi-modal* representations. Alternatively, they could be thought of as a "hub" that bind together modality-specific representation (e.g., Rogers & McClelland 2004), or, perhaps more precisely – merely as a "switchboard" between the modality-specific cortex, where the "real action" of semantic cognition takes place.

We construe multi-modal abstracta as entities whose concreta are modality-specific abstracta. As such, these abstracta indeed have "switchboard" capacities. However, we argue that they constitute a distinct representational layer – upon which categorical abstracta are instantiated. This distinction may seem to some as being irrelevant, but it is not. The difference between our view and the "switchboard" view is that we contend that the processing of non-perceptual information (e.g., "what is the capital of Norway?") does not *require* the involvement of the modality-specific cortex. A "widely-distributed, modality-specific view" of semantic cognition will predict the *mandatory* involvement of the modality-specific representations.

This nuance is critical. If an individual subscribes to the view that modality-specific information is *sometimes* activated during conceptual processing, this does not make him/her a proponent of the strong "widely-distributed, modality-specific view." Thus, the question that we need to ask is – What is the evidence that modality-specific representations are necessarily activated during the processing of non-perceptual concepts?

As noted by the commentators, there is ample evidence that processing concrete language activates modality-specific regions of the brain, which is consistent with the embodiment view. However, this does not provide evidence for *necessity*. In contrast, we think that the evidence *against* the necessity view is strong:

(1) As we note in the manuscript, whereas specific studies often find involvement of sensorimotor areas of the brain in language comprehension, meta-analyses of language comprehension suggest that the processing of meaning occurs *outside* of the sensorimotor cortex.

    We think that such a finding is difficult to reconcile with the view that sensorimotor representations should *always* take part in semantic processing. However, several commentators (e.g., **Davis et al.**; **Borghi & Tummolini**) argue that different concepts predominantly rely on different modalities (e.g., visual, auditory, and motor). In such a case, averaging neural activity across different concepts will "wash away" modality-specific activations. We agree that this is a viable criticism. Better evidence against "strong embodiment" can come from studies that examined the processing of concepts that involve specific modalities, for example, from research on processing of action verbs that involve motor and visual components.

(2) Even when one carefully looks at the literature that examined language processing of a specific modality, there is strong evidence *against* the idea of *necessity*. For example, in Gilead et al. (2013), we asked participants to read concrete and abstract sentences in the past, present, and future tense. The concrete sentences described manual actions that provided participants with both visual and motor information (e.g., "He opened the door"), whereas the abstract sentences did not involve such specific motor and visual representations (e.g., "She respected the decision"). The comparison of concrete and abstract sentences yielded significant activity in the regions identified in meta-analyses of concrete versus abstract sentence processing (Wang et al. 2010), namely, only in areas associated with the visual imagery such as the fusiform gyrus. More importantly, this activation was evident only in past- and present-tense sentences. For future tense sentences, there was no evidence of increased activity in the visual cortex. Our interpretation of this finding (and other similar findings, e.g., Aravena et al. 2012; 2014; Tomasino & Rumiati 2013; Tomasino et al. 2014) is that the involvement of visual representations in the semantic task is ancillary, rather than obligatory.

(3) If meta-analyses "wash away" modality-specific activations because of their heterogeneity, we should expect that a greater focus on a specific modality should allow for the detection of modality-specific activations. However, the meta-analysis of "action verb" processing (Watson et al. 2013) did not find evidence for activation in motor or premotor regions, as predicted by the embodiment view.

(4) One could argue that the lack of evidence for motor involvement in Watson et al.'s meta-analysis stems, again, from "washing away" the effect. For example, maybe some action verbs rely more on the hands, feet, or mouth, and a meta-analysis misses this. However, those studies that conducted the most anatomically precise analyses (e.g., used functional localizers to precisely pin-point motor areas) often failed to support the embodiment hypothesis (e.g., Postle et al. 2008).

(5) Although much of the evidence supportive of the embodied cognition view comes from functional neuroimaging studies, it could be argued that such studies are nonetheless limited, for example, because of their low temporal resolution and correlative nature. Indeed, there is causal evidence that suggests that individuals that suffered motor cortex lesions exhibit impairments in action semantics (e.g., Kemmerer et al. 2012). However, as noted by Kemmerer (2015) in studies such as Kemmerer et al. (2012), it is possible that the participants who had motor lesions failed action semantics tasks because they also had lesions in other regions of the left hemisphere crucial for semantic cognition (e.g., LIFG). A comprehensive study that examined the correlation between lesion location and impairment (Tarhan et al. 2015) did not find an association between motor areas and action semantics (but rather, localized such impairments to posterior temporal areas).

https://doi.org/10.1017/S0140525X19002000

Thus, although there is compelling evidence that the modality-specific cortex is *sometimes* involved in semantic processing (e.g., Huth et al. 2016), there is similarly compelling evidence that the sensorimotor cortex is *not always* involved in semantic processing. As such, we think that the strong embodiment view has been convincingly falsified. Instead, the activity of the default network likely reflects multi-modal processing that does not necessarily require modality-specific instantiation.

R5. Are categorical abstracta indeed distinct from multi-modal abstracta?

Another criticism concerning our view of conceptual cognition came from proponents of distributional semantics (i.e., **Davis et al.**; **Montefinese, Ambrosini, Visalli, & Vinson** (**Montefinese et al.**)). Namely, the idea that the meaning of words can be captured by keeping track of the spatiotemporal proximity of different lemmas. For example, the word "equality" is correlated in natural language with the words "liberty" and "inclusiveness." It has long been suggested that statistical learning algorithms should be able to use this ecological co-occurrence to capture the meaning of words, including highly intangible concepts such as equality.

Indeed, research into natural language processing using artificial neural networks has been able to use such spatiotemporal associations between lemmas to devise useful representations that allow computers to perform many linguistic tasks (e.g., respond to our commands and simulate conversations, identify taxonomic relationships, and even solve word analogies; e.g., Mikolov et al. 2013).

The success of these advances in machine learning requires that we seriously consider the possibility that the acquisition of word meaning in humans is likewise merely a matter of keeping track of statistical co-occurrence. Namely, that concepts such as "executive branch" and "ball" both rely on simple associative learning, and can be thought of as multi-modal abstracta, or, in the terminology of Peirce, indexical relationships. As suggested by **Montefinese et al.**, concepts such as "ball" may rely on spatiotemporal association between multi-modal representations (e.g., the sight of a ball and the sound of the word "ball"), and expressions such as "executive branch" may (mostly) rely on associations between different lemmas (e.g., "government" and "legislature").

Although this account is attractive, we think that there is clear evidence that this example of machine learning is not a good description of how humans learn the meaning of (at least some) concepts. For example, I can teach you the following concept: A *Kapara* is a ceremony where an individual takes a live chicken by their shoulder blades and swings it above one's head exactly three times; during this process, one's sins are transferred to the chicken, such that the person performing the ceremony is absolved.

Even if you were unaware of what a Kapara is, you now understand the meaning of this word. This acquisition of meaning did not rely on a lengthy process of teaching you a model of the co-occurrence of the word Kapara with the rest of the words in the English language. In fact, you only saw one instance of this word. In our terminology, you received the criterion of substitutability of this abstractum via symbol-based interaction (i.e., reading). This criterion is the algorithm that allows you to identify the Kapara ritual next time you see it, and to perform it yourself.

We believe that such "experiments" provide clear evidence that some abstracta do not rely on statistical co-occurrence, and do not rely on innate criteria of substitutability. We term these categorical abstracta. As we review in the manuscript, there is evidence that indeed, such categorical representations rely on neural activity in areas that extend beyond the multi-modal cortex.

R6. Neglected entities?

Whereas some commentators argue that there is redundancy in our representational ontology, others have suggested that there are representational entities we may have missed out on.

One potentially neglected construct is **Grafman**'s "Structured event complexes," which are schematic representations of events. To us, Grafman's structured event complexes resemble the construct of *script*, which we discuss in our ontology, and we did not manage to find any distinguishing features between the two constructs (indeed, in Knutson et al. 2004, the concept of structured event complex and script are discussed as being identical).

According to our account, scripts are amalgams of abstracta that (unlike memory for episodes) remain invariant across different particular situations. As such, scripts should rely on modality-specific and multi-modal abstracta to a lesser extent; instead, they should rely more on categorical abstracta. As we review in the manuscript, research that compares memory of specific events and scripts indeed shows that the latter is associated with activation in frontotemporal regions that supposedly subserve categorical abstracta.

Based on his extant empirical work, **Grafman** suggests that structured event complexes (which, as we note, may be thought of as scripts) rely on the prefrontal cortex (PFC). Although we agree that scripts rely on the PFC, the difference between our perspectives is in the interpretation of these findings. We suggest that the association between the PFC and scripts stems from the PFC's role in subserving categorical abstracta. Grafman argues that scripts are the basic representational building block that subserves PFC activity.

Evidence for our view comes from research showing that processing of relatively abstract objects (rather than events) is associated with PFC activity. For example, as we note in the manuscript the processing of superordinate objects (vs. basic level) concepts, and intangible (vs. tangible) categories recruits the PFC. Future research on the theory of structured event complexes may be able to accommodate such findings regarding the processing of non-event representations. Such attempts at theoretical arbitrations between our account and **Grafman**'s account may help scientists' understanding of the functionality of the PFC.

Popescu & Fitch's fascinating commentary opens up a crucial question regarding the representational capacities that subserve human generativity – and the dynamic generation of complex embedded structures – evident in human language and music.

We suggested the notion of *predicator* as the representational entity that subserves much of the generativity of cognition. Our discussion of the construct of predicator is grounded in linguistics; within our account, predicators endow us with "compositional semantics" in that one predicator modifies its argument in a predictable manner. **Popescu & Fitch** suggest that linguistic predicators are a special case of a broader representational group of "partially unsaturated" entities that they call "Treelets": "*These are subunits with open 'slots' that allow them to flexibly combine with other subunits in constrained ways.*"

These purported subunits can be observed, for example, in music: "just as a transitive verb has 'slots' demanding an animate subject and an object, a tonic–subdominant–dominant chord sequence allows a *constrained set of continuations*." Indeed, the

https://doi.org/10.1017/S0140525X19002000

tonic–subdominant–dominant (henceforth, TS…) subunit bears much resemblance to verbs that require an argument. Just like how the verb "John opened…" begs you to continue it with an object, the chord progression C major, followed by F major begs you for an answer of what comes next (typically, G major), and as such it is at least in some sense "unsaturated."

If the TS… subunit resembles the notion of predicator, what makes it distinct? **Popescu & Fitch** suggest that the TS… subunits are not predicators because, unlike predicators, they "make no semantic reference to the world, and entail no truth values." Indeed, according to our conceptualization, predicators refer – in the sense that they have concreta. Furthermore, they have a truth value – in the sense that they, similar to other types of abstracta, are an embodiment of beliefs about substitutability (and as such, similar to any belief, have a "mind-to-world direction of fit," and therefore, a truth value). However, in our account of abstraction, all of the representational building blocks of cognition (i.e., abstracta) can be thought of as having both sense and reference. Thus, we feel that it may be warranted to consider the TS… entity as a predicator nonetheless.

What we see as the main difference between TS… and predicators such as "…opened…" is that the latter are categorical abstracta (i.e., their criteria of substitutability are based on symbolic interaction) and the TS… predicator is a multi-modal abstractum (i.e., typically acquired via spatiotemporal association).

In the target article, we discuss predicators as a unique type of categorical abstractum. Indeed, those predicators that seem to be crucial for higher-order cognition and for a "language of thought" (e.g., logical predicators, mental state predicators, and mathematical predicators) are categorical abstracta, as their criteria of substitutability are *not* spatiotemporal. The case of musical subunits such as TS… reveals the possible existence of predicators on the level of spatiotemporal associations and even at the level of modality-specific abstractions. For example, an incomplete visual Gestalt can be thought of as a modality-specific predicator. The distinction between these types of predicators clearly deserves much attention in future behavioral and neurological studies.

R7. Constructivist critiques of our account of abstraction

We described prototypical acts of abstraction as forming the basis for mental travel, and as such, as preceding it. Several commentators have suggested that it is likely that mental travel often serves as the basis for acts of abstraction, and thus precedes it. This critique may be best summarized as such: we describe abstraction as a judgment, rather than a creative act. We relegate human creativity to the domain of mental travel.

Specifically, **Fedyk & Xu** suggest that human creativity does not find its proper place in our ontology and argue that it should be considered as one of the bases for acts of abstraction. In a similar vein, **Faber** hypothesizes that spontaneous processes of mental travel (i.e., mind wandering) may be functional in that they generate various creative permutations of events that enrich the storehouse of mental content. This synthesis of observations (which has proven to be a useful method to increase generalization of machine-learning models) can then be used as the basis for abstraction. The importance of "housekeeping" operations that occur whenever the mind is decoupled from the "here and now" is also highlighted by **Visalli & Cellini**, who discuss the role of sleep in the formation and updating of abstracta.

We agree with these comments that the act of mental travel (whether occurring during wake, sleep, or absent-mindedness) provides essential building blocks for acts of abstraction. As such, the relationship between abstraction and mental travel is likely a two-way street.

We discussed three routes through which we attain criteria of substitutability: innate, based on personal experiences, and based on symbol-based interaction with other humans. **Fedyk & Xu** argue that the contents of one's own mental travels should be considered a fourth origin. We think that this is a legitimate way to construe things. Nonetheless, it is worthwhile to recognize that there is a logical asymmetry between the three origins we discuss and this fourth one. Mental travel necessarily relies on abstraction, and abstraction relies on criteria of substitutability. Criteria of substitutability indeed *sometimes* originate from our mental travels – but are sometimes given to us (by nature or by our social environment) "for free."

Just like **Fedyk & Xu**, so do **Burman** and **Shillcock** worry that we have neglected the generative nature of human cognition in our model of abstraction. Specifically, Burman sees an affinity between our account and Piaget's early theorizing, but not with his later work. Piaget's better-known theory construes abstraction through the perspective of extensional logic, which means that representations are strictly defined by their referents or, in our terminology, their concreta. In his later studies, Piaget saw this formulation as a mistake and reconstrued mental representations as instantiating intensions/senses. Likewise, Shillcock perceives our description of abstraction as relying on purely inductive processes: "The authors' ontology reveals the conventional positivist notion of stepwise building up, from supposedly assumptionless 'atomic' foundations, verifying each move, until the highest processing is achieved." As such, he suggests that our account may apply to recognition tasks carried out by deep neural networks, but does not capture the productive agency of human cognition.

We think that this characterization of our view of abstraction as a bottom-up process that is based on the logic of extensions is a misperception of our perspective. Indeed, our account is described from the bottom up, reflecting the ontogenetic and phylogenetic emergence of higher-order cognition. However, we explicate that the abstractum is not merely a "grouping" of exemplars. Instead, the abstractum is a mental device that instantiates "criteria of substitutability," which is de facto, an intension. Moreover, the abstractum we describe is generative and is akin to speculative theory building, rather than a positivist collection of absolute facts. As we note: "criteria of substitutability can take a form such as 'things that are tasty are often made of chocolate'; this means that they can implement a theory (Murphy & Medin 1985). Theories allow us to generate predictions of (or imagine) future members of a set, rather than just assign a probability of class membership given a list of features (i.e., they implement a generative model, Ng & Jordan 2002; Helmholtz 1856)."

Indeed, as suggested by **Shillcock**, feed-forward neural networks are only able to recognize patterns, and do not capture the productivity of human thought; however, bi-directional neural networks, as assumed in the predictive cognition world, implement generative models – and begin to instantiate such productivity.

R8. Theory versus simulation and abstract versus concrete representation

We suggested two distinct routes by which mental travel occurs – simulation and theory-based inference. Furthermore, we suggested an affinity between simulation and relatively concrete representation and inference and abstract representation.

https://doi.org/10.1017/S0140525X19002000

Several commentators suggested some qualifications to this characterization.

Dohrn provides an in-depth discussion of the notion of simulation, relying on the rich philosophical history on the topic. He disagrees with our view that mental simulation entails an "autonoetic" component, and suggests that a mental image of a boulder rolling down a hill is also a simulation, but does not involve the presence of an experiencing self.

We believe that **Dohrn**'s response highlights that we have not adequately explicated our view of simulation. We think of the rolling boulder example highlighted by Dohrn is an example of vivid *mental imagery*. Our view is that one can engage in concrete imagery without engaging in a simulation – because, per (our) definition (following a similar definition by Goldman 2005) simulation is the process wherein a self-projected agent experiences the imagined world, and "reacts" to (simulated) occurrences.

Consider the example of a quarterback who is contemplating his next play. This quarterback may imagine the play from the perspective of a person sitting in the bleachers, or simulate his perspective as the actor holding the ball in his hands. Only if the quarterback "self-projects" into the field and allows his reality-oriented processes react to the event (i.e., "this huge guy is running at me! Ahh! I have to get rid of the ball!"), then we will say he engaged in simulation. The vivid representation of the field from the bleachers can highlight relationships between different states; the field becomes an abstract map of the situation – which is maintained in mind in a modality-specific manner. This 'bleachers view' is identical to how mathematical equations represent abstract ideas. The specifics of the font by which the mathematical formula is represented is irrelevant and only serves as a stand-in to convey highly intangible ideas; thus, such processes of mathematical reasoning in front of an imaginary blackboard are nonetheless theory-based inferences, supported by imagery (rather than simulation) processes.

According to our account, this vividness that simulation entails creates a difficulty in that it raises the danger of actual confusion between simulated and real events. For example, a person who vividly simulates taking a prescription pill might be confused and believe that the pill has already been taken. **Reyna & Broniatowski** argue that the opposite is true and that "concreteness and imageability are negatively related to false memory, per theory (Brainerd et al. 2008)." The findings Reyna & Broniatowski refer to are of higher levels of correct recall for concrete words in the DRM paradigm. Indeed, the concreteness of a word can improve its ability to be remembered. However, this does not entail that the concreteness of a simulation cannot lead to a confusion wherein a simulated event is perceived as real. Much research in the "imagination inflation" literature conclusively shows that this is indeed the case, and that simulation of an event can cause a blending between actually experienced events and imagined ones.

Whereas the imagination inflation literature highlights how simulation can permeate into one's representation of reality, **Mrkva, Cian, & Van Boven** (**Mrkva et al.**) suggest that the permeable boundaries of simulation and reality are bidirectional. Specifically, they highlight how the contents of one's simulations can be invaded by direct experience, limiting adaptive decision making. When I am contemplating whether to go on a trip to Italy, I may simulate walking the streets of Rome, eating pasta and drinking wine; the question of whether I have just eaten dinner or not should not matter for my decision processes. However, as explicated by Mrkva, Cian, and Van Boven, such fleeting sensations have substantial effects on one's decision making.

A prediction of our account, which is entailed by **Mrkva et al.**'s commentary, is that decisions that are based on theory-based inference should be less susceptible to one's experience in the here-and-now as compared with decisions that rely on simulation. This prediction can be readily tested. For example, such research can use the finding that prompting participants to simulate a future event in a concrete manner increases their propensity to save money (Peters & Büchel 2010); we predict that manipulations of one's current mood will have a greater effect on saving in the condition where participants are prompted to concretely simulate the future.

As noted, the concreteness of the simulation is an integral part of its usefulness in that it confuses us and react to the simulation as if it were the reality. However, we may not have been explicit enough in qualifying that this affinity between simulation-and-concrete, and theory-and-abstract is *statistical rather than absolute*. As noted by **Dohrn**, and discussed above, theory-based inference can sometimes be assisted by concrete representations (e.g., mathematical symbols on a blackboard). Furthermore, as highlighted by **Zhao, Mildner, & Tamir** (**Zhao et al.**), simulation often involves a complex interaction between abstract and concrete representations.

Zhao et al. establish their argument based on a fascinating study by Meyer et al. (2019). This study found that creative experts can construct vivid representations of psychologically distal scenarios by relying on activity in brain areas associated with high-level abstracta (e.g., dmPFC, angular gyrus, LIFG, and the temporal pole). At face value, the fact that the greater vividness of the representations created by creative experts relies on areas associated with categorical abstracta may seem inconsistent with our model.

However, as we discuss, one of the most important functionalities of highly abstract mental representations is in their ability to facilitate the creative generation of novel models via permutation and analogical reasoning. Our account specifically highlights the role of predicators in the creative process – as these entities subserve the ability for conceptual combination and analogical reasoning. Indeed, consistent with this view, in the research by Meyer et al., regions that supposedly subserve predication were those found to be essential for creative outputs.

Predicators are somewhat confusing – they are high-level abstracta that concretize a mental representation. When analyzing neural activity associated with predication, one should first and foremost expect to find the involvement of regions that are involved in high-level representation (e.g., angular gyrus and LIFG); however, as a consequence of this activity, we should also expect to see the involvement of lower-level representations.

Thus, we predict that when participants in Meyer et al.'s study produced vivid construals, activity in areas associated with high-level abstracta must have been correlated with activity in areas associated with vivid imagery. Namely, that the relationship between condition (creatives/controls) and activity in the medial temporal subsystem should be mediated by angular gyrus activity and moderated by vividness ratings. Such a prediction of our model could be readily examined in further analysis of the study by Meyer et al.

R9. The pragmatic perspectives

Our paper argues that conceptualizations within the predictive cognition world often endorse overly simple models of mental representations. We suggested that research into "model-based"

https://doi.org/10.1017/S0140525X19002000

reinforcement learning neglects the diversity of models/representations employed by cognition. Because of this limited view of representation, we think that this literature also neglects the role of discursive/deductive processes in decision making.

Dayan's commentary describes how Bayesian Decision Theory (BDT) may view the topics addressed in the paper. Dayan proposes that Bayesian decision theorists such as himself are "simple-minded sophisticates." Namely, scholars who use technically complex models but try to explain the brain and behavior by adhering to several straight-forward, common-sense principles. Their view is that individuals survive by making good decisions and developed brain mechanisms that support such decisions. To make good decisions, individuals characterize their current predicament, evaluate the consequences of the potential choices that are available in this predicament, and choose based on these evaluations.

According to **Dayan**, adopting such a view entails that the process of mental simulation should not be seen as fundamentally distinct from theory-based inference. Simulation does not entail reliance on life-like representations because the Bayesian Decision Maker does not need to predict actual future *events*. All that needs to be predicted is the *value* of such events. In the words of Dayan: "BDTs only need to predict evaluations – predicting in more detail what will happen is at most a means to this particular end"; "we only really need to predict evaluations rather than actual future outcomes."

We perceive an insufficiency in this focus on valuation as the driver of decision making. Surely, valuation is crucial, but we see it as the simplest aspect of the decision, one that merely entails crunching the numbers in the Bayesian computer. When predicting the value of actions, the challenge is to predict outcomes and imagine hypothetical realities. This part of cognition that BDTs may see as "at most a means toward an end" is what we see as the central preoccupation of the human mind.

Consider the example of an investment firm that tries to predict the *value* of its current portfolio in 3 years. This value may depend on the price of metal; the price of metal may depend on the question of who will be the American president in the next years; this question depends on whether Senators decide to remove the current president from office or not; this will be decided by factoring the senators' value system, political calculations, and so on.

These intermediary steps all rely on various models/representations and different processes, including processes of abstract deduction and vivid simulation. Surely, it is not the case that Bayesian Decision Models cannot integrate such complexities under their framework. However, the BDT perspective on human cognition often attempt to skirt around the messiness of such complex "inferential cranks" that are associated with model-based decisions.

Most notably, BDT typically construes model-based decision making as relying on variants of state-space search algorithms that operate upon episodic or semantic representations. The problem with this approach is that it assumes away the central challenge of the decision-maker. The state-space itself (or, in our words, the target representation) is typically not given to the problem solver and must be construed from a plethora of potential source representations.

Similarly to **Dayan**, **Spurrett** argues for the centrality of processes of evaluation in the predictive brain. Given the supposed importance of utility calculations for organisms' survival, Spurrett is surprised that we have not explicated how utility is represented in the brain and cognition. Specifically, we have not addressed a central question in Decision Theory regarding whether the brain employs an abstract "common currency" for utility calculations – or whether it uses concrete utility representations for different types of pleasures and pains.

Indeed, our paper does not address this question. However, the same principles we outlined with regard to the representation of actions, objects, and relationships translate easily to the representation of utility. Specifically, in our terminology, the question of common-currency can be recast as whether utility representations are modality-specific, multi-modal, or categorical.

We defined modality-specific abstracta as those that rely on "sensory impressions." However, a more accurate definition of "sensory" is – "that for which organisms have some innate detection mechanism." Organisms detect the sensory impression of "thermal," "chemical," and "mechanical" pain using different receptors. Some receptors are "polymodal" and respond to all three types of pain perturbations. According to our account, such polymodal receptors are still modality-specific, because they do not require any sort of association learning. Thus, pain may be effectively discussed as a single modality. Similarly, the sense of thirst stems from receptors that detect levels of osmolyte concentration and receptors that detect water volume, and are all part of a single system, and thereby, a modality.

It is possible that organisms have innate, polymodal "aversion" detectors that similarly respond to pain and thirst (without reliance on association learning). If such a system exists, then this representation of (negative) value serves as a *modality-specific common currency representation* and should exist outside of the multi-modal brain, potentially in sub-cortical areas.

Importantly, even if such modality-specific common currency representations indeed exist, they may live side-by-side with multi-modal and categorical common-currency representations. For example, representations of aversion can be associated with conditioned stimuli (e.g., the light that precedes the electric shock). This will give rise to a multi-modal abstractum that allows the organism to treat both the light and the shock as equivalent. Furthermore, some common currency representations of negative utility rely on categorical abstracta, acquired from society (e.g., the valuation of death as being a better outcome than the betrayal of one's sacred values). Such common-currency representations will likely be supported by frontotemporal areas that subserve categorical abstracta (rather than subcortical and modality-specific representations).

Thus, an implication of our view of abstraction and our pluralistic account of representations is that one should not expect to find a *single* "common currency" representation in the brain. Instead, one should expect to find various types of common currency representations, whose physiological bases depends on the extent to which they rely on modality-specific, multi-modal, or categorical abstracta.

The idea that representations of value can be modality-specific, multi-modal, or categorical has important ramifications for choice behavior. **Ledgerwood, Eastwick, & Gawronski**'s commentary explicates some of these consequences concerning the choices people make in social settings. They suggest that contextual factors that influence the level of construal determine whether individuals rely on categorical representations of value (which they term "ideas regarding liking") rather than on the concrete/embodied experience of value.

Importantly, there may be many cases where representations of value at different levels diverge. For example, a person eating

https://doi.org/10.1017/S0140525X19002000

at an over-hyped, expensive restaurant may have the abstract idea that she is having the time of her life, whereas her embodied representations may signal to her that this is an aversive experience. Thus, a representational diversity account of value may entail that one's construal of pain and pleasure can sometimes be quite fragmented. Such a multi-level account of value can have significant ramifications for people's construal of their (and others') choices.

R10. The social construction perspectives

The complexity of the human condition, entailed by our ability to generate and transfer categorical abstracta, extends well beyond the realm of valuation processes. This point is clearly articulated in the commentary by **Rossignac-Milon, Pinelli, & Higgins** concerning the social nature of abstraction and mental travel.

Their response highlights several aspects of the intersubjective aspects of mental travel, which we have not sufficiently highlighted. We discussed how meanings are transmitted from mind to mind; however, an important, understudied process is how two or more individuals jointly construct meaning. Such processes can sometimes be reduced to the workings of each mind, but sometimes may have emergent properties that may require their own explanatory mechanisms.

As we note, the creation of collective meaning is not merely a palatable experience but may have had actual survival benefits. We highlighted how intangible concepts could facilitate cohesion by generating facts that cannot be verified, and as such, can be used to generate myths that support social structure. **Montefinese et al.** extend this idea. They suggest that intangible constructs create interdependence on others – once we realize that we cannot figure out their meaning on our own. The uncertainty created by newly "discovered" intangibles pulls people together into communities where such puzzles are generated (i.e., intangible constructs are created) and then jointly solved. In other words, Montefinese et al. suggest that humans' fascination and preoccupation with abstract ideas is adaptive, in that it provides people with mysteries that bind them together.

As evident by the numerous fascinating responses to our paper, the sense of mystery surrounding concepts such as "abstraction" and "mental travel" can indeed prompt people to engage in constructive, joint efforts to comprehend the intangibles of the mind.

References

[The letters "a" and "r" before author's initials stand for target article and response references, respectively]

Adams, R., Stephan, K., Brown, H., Frith, C. & Friston, K. (2013) The computational anatomy of psychosis. *Frontiers in Psychiatry* **4**(47). Available at: http://doi.org/10.3389/fpsyt.2013.00047. [aMG]

Addis, D. R., Wong, A. T. & Schacter, D. L. (2007) Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia* **45**(7):1363–77. Available at: http://doi.org/10.1016/j.neuropsychologia.2006.10.016. [aMG]

Aglioti, S. M. & Pazzaglia, M. (2010) Representing actions through their sound. *Experimental Brain Research* **206**(2):141–51. doi: 10.1007/s00221-010-2344-x. [MP]

Aglioti, S. M. & Pazzaglia, M. (2011) Sounds and scents in (social) action. *Trends in Cognitive Sciences* **15**(2):47–55. doi: 10.1016/j.tics.2010.12.003. [MP]

Ajzen, I. & Fishbein, M. (1977) Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin* **84**:888–918. [AL]

Alba, J. W. & Hasher, L. (1983) Is memory schematic? *Psychological Bulletin* **93**(2):203–231. https://doi.org/10.1037/0033-2909.93.2.203. [VFR]

Alter, A. L. & Oppenheimer, D. M. (2009) Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review* **13**(3):219–35. [KM]

Altmann, G. T. M. (1997) *The ascent of Babel: An exploration of language, mind, and understanding*. OUP. [CPD]

Altmann, G. T. M. (2017) Abstraction and generalization in statistical learning: Implications for the relationship between semantic types and episodic tokens. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372** (1711):20160060. [CPD]

Altmann, G. T. M. & Ekves, Z. (2019) Events as intersecting object histories: A new theory of event representation. *Psychological Review* **126**(6):817–40. [CPD]

Amemiya, S. & Redish, A. D. (2016) Manipulating decisiveness in decision making: Effects of clonidine on hippocampal search strategies. *Journal of Neuroscience* **36** (3):814–27. Available at: http://doi.org/10.1523/jneurosci.2595-15.2016. [aMG]

Apperly, I. A. (2008) Beyond simulation-theory and theory-theory: Why social cognitive neuroscience should use its own concepts to study "theory of mind." *Cognition* **107** (1):266–83. Available at: http://doi.org/10.1016/j.cognition.2007.07.019. [aMG]

Aravena, P., Courson, M., Frak, V., Cheylus, A., Paulignan, Y., Deprez, V. & Nazir, T. (2014) Action relevance in linguistic context drives word-induced motor activity. *Frontiers in Human Neuroscience* **8**:163. [rMG]

Aravena, P., Delevoye-Turrell, Y., Deprez, V., Cheylus, A., Paulignan, Y., Frak, V. & Nazir, T. (2012) Grip force reveals the context sensitivity of language-induced motor activity during "action words" processing: Evidence from sentential negation. *PLOS ONE* **7**(12):e50287. [rMG]

Asch, S. E. (1951) Effects of group pressure upon the modification and distortion of judgment. In: *Groups, leadership and men*, ed. H. Guetzkow, pp. 177–90. Carnegie Press. [aMG]

Asimov, I. (1950) *I, Robot*. Gnome Press, New York. [aMG]

Attias, H. (2003) Planning by probabilistic inference. *Proceedings of the 9th International Workshop on Artificial Intelligence and Statistics*. [KFr]

Auksztulewicz, R., Schwiedrzik, C. M., Thesen, T., Doyle, W., Devinsky, O., Nobre, A. C., Schroeder, C. E., Friston, K. J. & Melloni, L. (2018) Not all predictions are equal: "What" and "when" predictions modulate activity in auditory cortex through different mechanisms. *Journal of Neuroscience* **38**(40):8680–93. [MGV]

Austin, J. L. (1975) *How to do things with words*. Oxford university press. [aMG]

Babkoff, H., Sing, H. C., Thorne, D. R., Genser, S. G. & Hegge, F. W. (1989) Perceptual-distortions and hallucinations reported during the course of sleep-deprivation. *Perceptual and Motor Skills* **68**(3):787–98. Available at: http://doi.org/10.2466/pms.1989.68.3.787. [aMG]

Back, A. (2014) *Aristotle's theory of abstraction*. Springer. [AD]

Badre, D. & D'Esposito, M. (2007) Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience* **19** (12):2082–99. [aMG]

Badre, D. & D'Esposito, M. (2009) Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience* **10**(9):659–69. Available at: http://doi.org/10.1038/nrn2667. [aMG]

Badre, D., Kayser, A. S. & D'Esposito, M. (2010) Frontal cortex and the discovery of abstract action rules. *Neuron* **66**(2):315–26. Available at: http://doi.org/10.1016/j.neuron.2010.03.025. [aMG]

Badre, D. & Nee, D. E. (2018) Frontal cortex and the hierarchical control of behavior. *Trends in Cognitive Sciences* **22**(2):170–88. doi: doi:10.1016/j.tics.2017.11.005. [JG]

Baetens, K., Ma, N., Steen, J. & Overwalle, F. V. (2014) Involvement of the mentalizing network in social and non-social high construal. *Social Cognitive and Affective Neuroscience* **9**(6):817–24. https://doi.org/10.1093/scan/nst048. [ZZ]

Baillargeon, R., Spelke, E. S. & Wasserman, S. (1985) Object permanence in five-month-old infants. *Cognition* **20**(3):191–208. [aMG]

Baker, C. L. & Tenenbaum, J. B. (2014) Modeling human plan recognition using Bayesian theory of mind. In: *Plan, activity, and intent recognition*, ed. G. Sukthankar, C. Geib, H. H. Bui, D. V. Pynadath & R. P. Goldman, pp. 177–204. Morgan Kaufmann. [KFr]

Baldassano, C., Hasson, U. & Norman, K. A. (2018) Representation of real-world event schemas during narrative perception. *Journal of Neuroscience* **38**(45):9689–99. https://doi.org/10.1523/JNEUROSCI.0251-18.2018. [MlS]

Baldeweg, T. (2006) Repetition effects to sounds: Evidence for predictive coding in the auditory system. *Trends in Cognitive Sciences* **10**(3):93–94. [MtS]

Ballesio, A. & Cellini, N. (2019) Updating internal cognitive models during sleep. *Journal of Neuroscience* **39**(11):1966–68. [AV]

Banaji, M. R., Hardin, C. & Rothman, A. J. (1993) Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology* **65**(2):272. [aMG]

Bar, M. (2007) The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences* **11**(7):280–89. [LSP]

Bar, M., Aminoff, E., Mason, M. & Fenske, M. (2007) The units of thought. *Hippocampus* **17**(6):420–28. https://doi.org/10.1002/hipo.20287. [ZZ]

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmidt, A. M., Dale, A. M., Hämäläinen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R. & Halgren, E. (2006) Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America* **103**(2):449–54. Available at: http://doi.org/10.1073/pnas.0507062103. [aMG]

Bar, M. (Ed.). (2011) *Predictions in the brain: Using our past to generate a future*. Oxford University Press. [aMG]

Barca, L., Mazzuca, C. & Borghi, A. M. (2017) Pacifier overuse and conceptual relations of abstract and emotional concepts. *Frontiers in Psychology* **8**:2014. [AMB]

Barca, L., Mazzuca, C. & Borghi, A. M. (in press) Overusing the pacifier during infancy sets a footprint on abstract words processing. *Journal of Child Language*. [AMB]

Bargh, J. A. (2006) What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology* **36**(2): 147–68. Available at: http://doi.org/10.1002/ejsp.336. [aMG]

Baron-Cohen, S. (1995) *Mindblindness: An essay on autism and theory of mind*. MIT Press. [MKB]

Baron-Cohen, S., Ashwin, E., Ashwin, C., Tavassoli, T. & Chakrabarti, B. (2009) Talent in autism: Hypersystemizing, hyper-attention to detail and sensory hypersensitivity. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **364** (1522):1377–83. [MKB]

Baron-Cohen, S., Knickmeyer, R. C. & Belmonte, M. K. (2005) Sex differences in the brain: Implications for explaining autism. *Science* **310**(5749):819–23. [MB, MKB]

Baron-Cohen, S., Leslie, A. M. & Frith, U. (1985) Does the autistic child have a "theory of mind"? *Cognition* **21**(1):37–46. [aMG]

Barrett, L. F., Quigley, K. S. & Hamilton, P. (2016) An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B* **371** (1708): 20160011. [aMG]

Barsalou, L. W. (1983) Ad hoc categories. *Memory & Cognition* **11**(3):211–27. [AMB, aMG]

Barsalou, L. W. (1985) Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **11**(4):629. [AMB]

Barsalou, L. W. (1999) Perceptual symbol systems. *Behavioral and Brain Sciences* **22** (4):577–660. [aMG, KM]

Barsalou, L. W. (2008) Grounded cognition. *Annual Review of Psychology* **59**:617–45. [aMG]

Barsalou, L. W. (2009) Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364** (1521):1281–89. [aMG]

Bassok, M., Dunbar, K. N. & Holyoak, K. J. (2012) Introduction to the special section on the neural substrate of analogical reasoning and metaphor comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **38**(2):261. [aMG]

Bechara, A., Damasio, H. & Damasio, A. R. (2003) Role of the amygdala in decision-making. In: *Amygdala in brain function: Basic and clinical approaches, vol. 985*, ed. P. Shinnick-Gallagher, A. Pitkanen, A. Shekhar & L. Cahill, pp. 356-69. Academy of Sciences. [aMG]

Beilin, H. (1992) Piaget's new theory. In: *Piaget's theory: Prospects and possibilities*, ed. H. Beilin & P. B. Pufall, pp. 1–17. L. Erlbaum. [JTB]

Beker, S., Foxe, J. J. & Molholm, S. (2018) Ripe for solution: Delayed development of multisensory processing in autism and its remediation. *Neuroscience and Biobehavioral Reviews* **84**:182–92. [MKB]

Bélanger, M. J., Atance, C. M., Varghese, A. L., Nguyen, V. & Vendetti, C. (2014) What will I like best when I'm all grown up? Preschoolers' understanding of future preferences. *Child Development* **85**(6):2419–31. doi: 10.1111/cdev.12282. [BO]

Belmonte, M. K. (2008a) Does the experimental scientist have a 'theory of mind'? *Review of General Psychology* **12**(2):192–204. [MKB]

Belmonte, M. K. (2008b) Human, but more so: What the autistic brain tells us about the process of narrative. In: *Autism and representation*, ed. M. Osteen, pp. 166–79. Routledge. [MKB]

Bergen, B. & Chang, N. (2005) Embodied construction grammar in simulation-based language understanding. In: *Construction grammar(s): Cognitive and cross-language dimensions*, ed. J.-O. Östman & M. Fried, pp. 147–90. John Benjamins. [MlS]

Berger, J. (1985) *Statistical decision theory and Bayesian analysis*. Springer. [PD]

Bergmann, J., Morgan, A. T. & Muckli, L. (2019) Two distinct feedback codes in V1 for 'real' and 'imaginary' internal experiences. *BioRxiv* **664870**. [LSP]

Berkeley, G. (1710/1975) *The principles of human knowledge*. Everyman. (Original work published in 1710) [DD]

Berwick, R. C., Friederici, A. D., Chomsky, N. & Bolhuis, J. J. (2013) Evolution, brain, and the nature of language. *Trends in cognitive sciences* **17**(2):89–98. [aMG]

Bestgen, Y. & Vincze, N. (2012) Checking and bootstrapping lexical norms by means of word similarity indexes. *Behavior Research Methods* **44**:998–1006. [MM]

Bhatia, S., Richie, R. & Zou, W. (2019) Distributed semantic representations for modeling human judgment. *Current Opinion in Behavioral Sciences* **29**:31–36. [MM]

Bhatia, S. & Walasek, L. (2016) Event construal and temporal distance in natural language. *Cognition* **152**:1–8. Available at: http://doi.org/10.1016/j.cognition.2016.03.011. [aMG]

Binder, J. R. & Desai, R. H. (2011) The neurobiology of semantic memory. *Trends in Cognitive Sciences* **15**(11):527–36. doi: https://doi.org/10.1016/j.tics.2011.10.001. [aMG, MlS]

Binder, J. R., Desai, R. H., Graves, W. W. & Conant, L. L. (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex* **19**(12):2767–96. https://doi.org/10.1093/cercor/bhp055. [aMG, CPD, ZZ]

Blalock, G., Kadiyali, V. & Simon, D. H. (2009) Driving fatalities after 9/11: A hidden cost of terrorism. *Applied Economics* **41**(14):1717–29. [KM]

Bless, H., Clore, G. L., Schwarz, N., Golisano, V., Rabe, C. & Wölk, M. (1996) Mood and the use of scripts: Does a happy mood really lead to mindlessness? *Journal of Personality and Social Psychology* **71**(4):665–79. [KM]

Bloom, P. & Markson, L. (1998) Capacities underlying word learning. *Trends in Cognitive Sciences* **2**(2):67–73. Available at: http://doi.org/10.1016/s1364-6613(98)01121-8. [aMG]

Boduroglu, A., Shah, P. & Nisbett, R. E. (2009) Cultural differences in allocation of attention in visual information processing. *Journal of Cross-Cultural Psychology* **40** (3):349–60. [MKB]

Bogdan, R. J. (2009) *Predicative minds: The social ontogeny of propositional thinking*. Cambrdige, MA: MIT Press. [aMG]

Bond, R. & Smith, P. B. (1996) Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin* **119**(1):111–37. Available at: http://doi.org/10.1037/0033-2909.119.1.111. [aMG]

Borghi, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G. & Tummolini, L. (2019a) Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of Life Reviews* **29**:120–53. [AMB]

Borghi, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G. & Tummolini, L. (2019b) Words as social tools: Flexibility, situatedness, language and sociality in abstract concepts: Reply to comments on "Words as social tools: Language, sociality and inner grounding in abstract concepts." *Physics of Life Reviews* **29**:178–84. [AMB, MP]

Borghi, A. M., Barca, L., Binkofski, F. & Tummolini, L. (2018a) Abstract concepts, language and sociality: From acquisition to inner speech. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**(1752):20170134. [AMB]

Borghi, A. M., Barca, L., Binkofski, F. & Tummolini, L. (2018b) Varieties of abstract concepts: Development, use and representation in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**(1752):20170121. [AMB]

Borghi, A. M. & Binkofski, F. (2014) *Words as social tools: An embodied view on abstract concepts*. Springer. [AMB]

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C. & Tummolini, L. (2017) The challenge of abstract concepts. *Psychological Bulletin* **143**(3):263–92. [AMB]

Borghi, A. M., Caramelli, N. & Setti, A. (2005) Conceptual information on objects' locations. *Brain and Language* **93**(2):140–51. [AMB]

Borghi, A. M. & Zarcone, E. (2016) Grounding abstractness: Abstract concepts and the activation of the mouth. *Frontiers in Psychology* **7**:1498. [AMB]

Bornstein, A. M., Khaw, M. W., Shohamy, D. & Daw, N. D. (2017) Reminders of past choices bias decisions for reward in humans. *Nature Communications* **8**:15958. https://doi.org/10.1038/ncomms15958. [ZZ]

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D. (2001) Conflict monitoring and cognitive control. *Psychological Review* **108**(3):624–52. Available at: http://doi.org/10.1037//0033-295x.108.3.624. [aMG]

Boulenger, V., Hauk, O. & Pulvermüller, F. (2008) Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension. *Cerebral Cortex* **19**(8):1905–14. https://doi.org/10.1093/cercor/bhn217. [MlS]

Bower, G. H., Clark, M. C., Lesgold, A. M. & Winzenz, D. (1969) Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior* **8**(3):323–43. [aMG]

Bowers, J. S. (2009) On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review* **116** (1):220-51. Available at: http://doi.org/10.1037/a0014462. [aMG]

Bowman-Smith, C. K., Shtulman, A. & Friedman, O. (2019) Distant lands make for distant possibilities: Children view improbable events as more possible in far-away locations. *Developmental Psychology* **55**:722–28. doi: 10.1037/dev0000661. [BO]

Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L. & Mills, B. A. (2008) Semantic processing in "associative" false memory. *Psychonomic Bulletin & Review* **15** (6):1035–1053. doi: 10.3758/PBR.15.6.1035. [rMG, VFR]

Brandwein, A. B., Foxe, J. J., Butler, J. S., Frey, H. P., Bates, J. C., Shulman, L. H. & Molholm, S. (2015) Neurophysiological indices of atypical auditory processing and multisensory integration are associated with symptom severity in autism. *Journal of Autism and Developmental Disorders* **45**(1):230–44. [MKB]

Brandwein, A. B., Foxe, J. J., Butler, J. S., Russo, N. N., Altschuler, T. S., Gomes, H. & Molholm, S. (2013) The development of multisensory integration in high-functioning autism: High-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. *Cerebral Cortex* **23**(6):1329–41. [MKB]

Brentano, F. (1874) *Psychology from an empirical standpoint*. Routledge. 1995 (Original work published in 1874) [aMG]

Broniatowski, D. A. & Reyna, V. F. (2018) A formal model of fuzzy-trace theory: Variations on framing effects and the Allais Paradox. *Decision* **5**(4):205–52. https://doi.org/10.1037/dec0000083. [VFR]

Brooks, J. A., Chikazoe, J., Sadato, N. & Freeman, J. B. (2019) The neural representation of facial-emotion categories reflects conceptual structure. *Proceedings of the National*

Academy of Sciences **116**(32):15861–70. https://doi.org/10.1073/pnas.1816408116. [ZZ]
Brooks, J. A. & Freeman, J. B. (2018) Conceptual knowledge predicts the representational structure of facial emotion perception. *Nature Human Behaviour* **2**(8):581–91. https://doi.org/10.1038/s41562-018-0376-6. [ZZ]
Brown, R. (1958) How shall a thing be called? *Psychological Review* **65**(1):14. [MR-M]
Bruner, J. S. (1960) *The process of education. [A summary and synthesis of the Woods Hole conference of 1959]*. Harvard University Press. [JTB]
Buckner, R. L. & Carroll, D. C. (2007) Self-projection and the brain. *Trends in Cognitive Sciences* **11**(2):49–57. [aMG]
Buehler, R. & Griffin, D. (2003) Planning, personality, and prediction: The role of future focus in optimistic time predictions. *Organizational Behavior and Human Decision Processes* **92**(1–2):80–90. [KM]
Buehler, R., Griffin, D. & Ross, M. (1994) Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology* **67**(3):366–81. [KM]
Buehler, R., Griffin, D. W. & Ross, M. (2002) The triumph of hope over experience: Exploring the planning fallacy. In: *Heuristics and biases: The psychology of intuitive judgment*, ed. T. Gilovich, D. W. Griffin & D. Kahneman. Cambridge University Press. [KM]
Bunge, S. A., Wendelken, C., Badre, D. & Wagner, A. D. (2004) Analogical reasoning and prefrontal cortex: Evidence for separable retrieval and integration mechanisms. *Cerebral Cortex* **15**(3):239–49. [aMG]
Burgoon, E. M., Henderson, M. D. & Markman, A. B. (2013) There are many ways to see the forest for the trees: A tour guide for abstraction. *Perspectives on Psychological Science* **8**:501–520. [AD]
Burman, J. T. (2013) Updating the Baldwin effect: The biological levels behind Piaget's new theory. *New Ideas in Psychology* **31**(3):363–73. doi: 10.1016/j.newideapsych.2012.07.003. [JTB]
Burman, J. T. (2015) Neglect of the foreign invisible: Historiography and the navigation of conflicting sensibilities. *History of Psychology* **18**(2):146–69. doi: 10.1037/a0039194. [JTB]
Burman, J. T. (2016) Jean Piaget's neo-Gödelian turn: Between biology and logic, origins of the new theory. *Theory & Psychology* **26**(6):751–72. doi: 10.1177/0959354316672595. [JTB]
Burman, J. T. (2017) Philosophical histories can be contextual without being sociological: Comment on Araujo's historiography. *Theory & Psychology* **27**(1):117–25. doi: 10.1177/0959354316682862. [JTB]
Burman, J. T. (2019) Development. [originally entitled "Development under evolution"]. In: *The Cambridge handbook of the intellectual history of psychology*, ed. R. J. Sternberg & W. Pickren, pp. 287–317. Cambridge University Press. [JTB]
Burman, J. T. (in press) On Kuhn's case, and Piaget's: A critical two-sited hauntology (or, on impact without reference). [In: F. Callard & C. Millard, eds, Special issue dedicated to the memory of John Forrester, 1949–2015]. *History of the human sciences*. doi: 10.1177/0952695120911576 [JTB]
Buzsaki, G. (2010) Neural syntax: Cell assemblies, synapsembles, and readers. *Neuron* **68**. https://doi.org/10.1016/j.neuron.2010.09.023. [AD]
Camerer, C. F., Ho, T. H. & Chong, J. K. (2002) Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *Journal of Economic Theory* **104**(1):137–88. Available at: http://doi.org/10.1006/jeth.2002.2927. [aMG]
Campbell, J (1968) *The masks of God: Creative mythology*. New York: Viking Press. [aMG]
Campbell, J. (1991/1959) *The masks of God: Primitive mythology*. Arkana. [aMG]
Campbell, R. L. (2001) Reflecting abstraction in context. In: *Studies in reflecting abstraction*, ed. J. Piaget, pp. 1–27. Taylor & Francis. [JTB]
Campbell, R. L. (2009) Constructive processes: Abstraction, generalization, and dialectics. In: *The Cambridge companion to Piaget*, ed. U. Müller, J. I. M. Carpendale & L. Smith, pp. 150–70. Cambridge University Press. [JTB]
Carey, S. (2009) *The origin of concepts*. Oxford University Press. [aMG]
Carruthers, P. (2002) The cognitive functions of language. *Behavioral and Brain Sciences* **25**(06):657–74. [aMG]
Carver, C. S. & Scheier, M. F. (2001) *On the self-regulation of behavior*. Cambridge University Press. [aMG]
Cattell, R. B. & Horn, J. L. (1978) A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement* **15**:139–64. [AD]
Cavalli-Sforza, L. L. & Feldman, M. W. (1981) *Cultural transmission and evolution. A quantitative approach*. Princeton University Press. [aMG]
Chalmers, D. (2002) Does conceivability entail possibility? In: *Conceivability and possibility*, ed. T. S. Gendler & J. Hawthorne, pp. 71–125. Clarendon Press. [DD]
Chapman, L. J. & Chapman, J. P. (1978) The measurement of differential deficit. *Journal of Psychiatric Research* **14**(1–4):303–311. doi: 10.1016/0022-3956(78)90034-1. [JG]
Chapman, L. J. & Chapman, J. P. (2001) Commentary on two articles concerning generalized and specific cognitive deficits. *Journal of Abnormal Psychology* **110**(1):31–39. doi: 10.1037//0021-843x.110.1.31. [JG]
Chater, N. (2018) *The mind is flat*. Allen Lane. [CF]
Chater, N. & Vitanyi, P. (2003) Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences* **7**(1):19–22. Available at: http://doi.org/10.1016/s1364-6613(02)00005-0. [aMG]
Chen, Y. C. & Spence, C. (2010) When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition* **114**(3):389–404. doi: 10.1016/j.cognition.2009.10.012. [MP]
Chomsky, N. & Halle, M. (1968) *The sound pattern of English*. Harper and Row. [MtS]
Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N. & Andrews-Hanna, J. R. (2016) Mind-wandering as spontaneous thought: A dynamic framework. *Nature Reviews Neuroscience* **17**(11):718–31. https://doi.org/10.1038/nrn.2016.113. [MFa]
Chwilla, D. J. & Kolk, H. H. J. (2005) Accessing world knowledge: Evidence from N400 and reaction time priming. *Cognitive Brain Research* **25**:589–606. [MlS]
Cian, L. (2017) Verticality and conceptual metaphors: A systematic review. *Journal of the Association for Consumer Research* **2**(4):444–59. [KM]
Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* **36**(3):181–204. [arMG, AV, KFr]
Clark, J. E., Watson, S. & Friston, K. J. (2018) What is mood? A computational perspective. *Psychological Medicine* **48**(14):2277–84. Available at: http://doi.org/10.1017/s0033291718000430. [aMG]
Clayton, N. S., Dally, J. M. & Emery, N. J. (2007) Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist. *Philosophical Transactions of the Royal Society B-Biological Sciences* **362**(1480):507–22. [aMG]. Available at: http://doi.org/10.1098/rstb.2006.1992. [aMG]
Clements, G. N. (1985) The geometry of phonological features. *Phonology Yearbook* **2**:225–53. [MtS]
Cohen, G. (2000) Hierarchical models in cognition: Do they have psychological reality? *European Journal of cognitive psychology* **12**(1):1–36. [aMG]
Collins, A. G. E. & Frank, M. J. (2012) How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience* **35**:1024–35. [PD]
Collins, A. M. & Quillian, M. R. (1969) Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior* **8**(2):240–7. [aMG]
Connell, L., Lynott, D. & Banks, B. (2018) Interoception: The forgotten modality in perceptual grounding of abstract and concrete concepts. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **373**(1752):20170143. doi: 10.1098/rstb.2017.0143. [MM, MP]
Conson, M., Mazzarella, E., Esposito, D., Grossi, D., Marino, N., Massagli, A. & Frolli, A. (2015) "Put myself into your place": Embodied simulation and perspective taking in autism spectrum disorders. *Autism Research* **8**(4):454–66. [MKB]
Conway, M. A. & Pleydell-Pearce, C. W. (2000) The construction of autobiographical memories in the self-memory system. *Psychological Review* **107**(2):261–88. Available at: http://doi.org/10.1037//0033-295x.107.2.261. [aMG, ZZ]
Cooley, E. & Payne, B. K. (2019) A group is more than the average of its parts: Why existing stereotypes are applied more to the same individuals when viewed in groups than when viewed alone. *Group Processes & Intergroup Relations* **22**:673–87. [AL]
Corballis, M. C. (2014) *The recursive mind: The origins of human language, thought, and civilization*. Princeton University Press. [aMG]
Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K. & Powers, A. R. (2019) Hallucinations and strong priors. *Trends in Cognitive Sciences* **23**(2):114–27. https://doi.org/10.1016/j.tics.2018.12.001. [PL]
Coughlin, C., Robins, R. W. & Ghetti, S. (2019) Development of episodic prospection: Factors underlying improvements in middle and late childhood. *Child Development* **90**(4):1109–22. doi: 10.1111/cdev.13001. [BO]
Courchesne, E. & Allen, G. (1997) Prediction and preparation, fundamental functions of the cerebellum. *Learning & Memory* **4**:1–35. [MKB]
Craig, A. D. (2010) The sentient self. *Brain Structure and Function* **214**:563–577. [CF]
Crutch, S. J., Troche, J., Reilly, J. & Ridgway, G. R. (2013) Abstract conceptual feature ratings: The role of emotion, magnitude, and other cognitive domains in the organization of abstract conceptual knowledge. *Frontiers in Human Neuroscience* **7**:186. [MM]
Crutch, S. J. & Warrington, E. K. (2004) Abstract and concrete concepts have structurally different representational frameworks. *Brain* **128**(3):615–27. [MM]
Currie, G. (1995) *Image and mind: Film, philosophy, and cognitive science*. Cambridge University Press. [DD]
Cutler, A. (2017) *The cornerstones of twenty-first century psycholinguistics*. https://doi.org/10.4324/9781315084503-1. *[TP]*
d'Acremont, M., Fornari, E. & Bossaerts, P. (2013) Activity in inferior parietal and medial prefrontal cortex signals the accumulation of evidence in a probability learning task. *PLOS Computational Biology* **9**:e1002895. doi: 10.1371/journal.pcbi.1002895. [VFR]
D'Argembeau, A. & Mathy, A. (2011) Tracking the construction of episodic future thoughts. *Journal of Experimental Psychology. General* **140**(2):258–71. https://doi.org/10.1037/a0022581. [ZZ]
Danker, J. F., Tompary, A. & Davachi, L. (2016) Trial-by-trial hippocampal encoding activation predicts the fidelity of cortical reinstatement during subsequent retrieval. *Cerebral Cortex* **27**(7):3515–24. [aMG]

Davachi, L. (2006) Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology* **16**(6):693–700. [aMG]
Davidson, P. M. (1988) Piaget's category-theoretic interpretation of cognitive development: A neglected contribution. *Human Development* **31**(4):225–44. doi: 10.1159/000275811. [JTB]
Davidson, P. M. (1993) Commentary [in response to R. B. Ricco's "Revising the logic of operations as a relevance logic"]. *Human Development* **36**(3):147–49. doi: 10.1159/000277332. [JTB]
Davis, C. P., Altmann, G. & Yee, E. (2020) Situational systematicity: A role for schema in understanding the differences between abstract and concrete concepts. *Cognitive Neuropsychology*. https://doi.org/10.1080/02643294.2019.1710124. [CPD]
Davis, C. P., Joergensen, G. H., Boddy, P., Dowling, C. & Yee, E. (in press) Making it harder to "see" meaning: The more you see something, the more its conceptual representation is susceptible to visual interference. *Psychological Science*. [CPD]
Davis, C. P. & Yee, E. (2019) Features, labels, space, and time: Factors supporting taxonomic relationships in the anterior temporal lobe and thematic relationships in the angular gyrus. *Language, Cognition and Neuroscience* **34**(10):1347–57. [CPD]
Davison, A. J. & Murray, D. W. (2002) Simultaneous localization and map-building using active vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(7):865–80. doi: 10.1109/tpami.2002.1017615. [KFr]
Daw, N. D. & Dayan, P. (2014) The algorithmic anatomy of model-based evaluation. *Philosophical Transactions of the Royal Society B-Biological Sciences* **369**(1655):12. Available at: http://doi.org/10.1098/rstb.2013.0478. [aMG]
Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**(6):1204–15. Available at: http://doi.org/10.1016/j.neuron.2011.02.027. [aMG]
Daw, N. D., Niv, Y. & Dayan, P. (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience* **8**(12):1704–711. [PD]
Dayan, P. (1993) Improving generalization for temporal difference learning: The successor representation. *Neural Computation* **5**:613–24. [PD]
Deacon, T. (1997) *The symbolic species*. Norton. [aMG]
Dehaene, S. (2003) The neural basis of the Weber–Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences* **7**(4):145–47. [KFi]
Dehaene, S., Piazza, M., Pinel, P. & Cohen, L. (2003) Three parietal circuits for number processing. *Cognitive Neuropsychology* **20**(3-6):487–506. Available at: http://doi.org/10.1080/02643290244000239. [aMG]
de Heer, M. & Papathomas, T. V. (2017) The Ames-window illusion and its variations. In: *The Oxford compendium of visual illusions*, ed. A. Shapiro & D. Todorovic. [JBD]
DeLoache, J. S. (2004) Becoming symbol-minded. *Trends in Cognitive Sciences* **8**(2):66–270. [aMG]
De Luca, F., Benuzzi, F., Bertossi, E., Braghittoni, D., di Pellegrino, G. & Ciaramelli, E. (2018) Episodic future thinking and future-based decision-making in a case of retrograde amnesia. *Neuropsychologia* **110**:92–103. Available at: http://doi.org/10.1016/j.neuropsychologia.2017.08.007. [aMG]
Demetriou, A., Makris, N., Kazi, S., Spanoudis, G. & Shayer, M. (2018a) The developmental trinity of mind: Cognizance, executive control, and reasoning. *WIREs Cognitive Science* **2018**:e1461. https://doi.org/10.1002/wcs.1461. [AD]
Demetriou, A., Makris, N., Kazi, S., Spanoudis, G., Shayer, M. & Kazali, E. (2018b) Mapping the dimensions of general intelligence: An integrated differential-developmental theory. *Human Development* **61**:4–42. http://dx.doi.org/10.1159%2F000484450. [AD]
Demetriou, A. & Spanoudis, G. (2018) *Growing minds: A general theory of intelligence and learning*. Routledge. [AD]
Denève, S. & Jardri, R. (2016) Circular inference: Mistaken belief, misplaced trust. *Current Opinion in Behavioral Sciences* **11**:40–48. https://doi.org/10.1016/j.cobeha.2016.04.001. [PL]
Dennett, D. & Haugeland, J. (1987) Intentionality. In: *The Oxford companion to the mind*, ed. R. L. Gregory. Oxford University Press. [aMG]
Deręgowski, J. B. (2014) Perspectival reversals of a solid: A phenomenal description. *Perception* **43**:1261–63. [JBD]
Deręgowski, J. B. (2017) Cross-cultural studies of illusions. In: *The Oxford compendium of visual illusions*, ed. A. Shapiro & D. Todorovic, pp. 174–84. Oxford University Press. [JBD]
Desai, R. H., Conant, L. L., Binder, J. R., Park, H. & Seidenberg, M. S. (2013) A piece of the action: Modulation of sensory-motor regions by action idioms and metaphors. *NeuroImage* **83**:862–69. https://doi.org/10.1016/j.neuroimage.2013.07.044. [MlS]
Desai, R. H., Reilly, M. & van Dam, W. (2018) The multifaceted abstract brain. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**(1752):20170122. https://doi.org/10.1098/rstb.2017.0122. [AMB, MlS]
Desimone, R., Albright, T. D., Gross, C. G. & Bruce, C. (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *The Journal of Neuroscience* **4**(8):2051–62. [aMG]
Dickinson, A. (1985) Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London B, Biological Sciences* **308**(1135):67–78. [aMG]
Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M. A. & Friston, K. (2017) Distinct top-down and bottom-up brain connectivity during visual perception and imagery. *Scientific Reports* **7**, Article number: 5677. [LSP]
Dohmatob, E., Dumas, G. & Bzdok, D. (2018) Dark control: Towards a unified account of default mode function by Markov decision processes. *bioRxiv* **148890**. [LSP]
Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D. & Daw, N. D. (2015) Model-based choices involve prospective neural activity. *Nature Neuroscience* **18**(5):767–72. Available at: http://doi.org/10.1038/nn.3981. [aMG]
Doumas, L. A. A., Hummel, J. E. & Sandhofer, C. M. (2008) A theory of the discovery and predication of relational concepts. *Psychological Review* **115**(1):1–43. Available at: http://doi.org/10.1037/0033-295x.115.1.1. [aMG]
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. [AMB]
Doya, K. (2007) *Bayesian brain: Probabilistic approaches to neural coding*. MIT press. [KFr, rMG]
Dreyer, F. R. & Pulvermüller, F. (2018) Abstract semantics in the motor system? – An event-related fMRI study on passive reading of semantic word categories carrying abstract emotional and mental meaning. *Cortex* **100**:52–70. [AMB]
Ducret, J.-J. (1988) Operatory logic revisited (A. Cornu-Wells, trans.). *New Ideas in Psychology* **6**(3):357–69. doi: 10.1016/0732-118X(88)90048-7. [JTB]
Dumais, K. M., Chernyak, S., Nickerson, L. D. & Janes, A. C. (2018) Sex differences in default mode and dorsal attention network engagement. *PLOS ONE* **13**(6):e0199049. [MKB]
Dunbar, R. & Dunbar, R. I. M. (1998) *Grooming, gossip, and the evolution of language*. Harvard University Press. [aMG]
DuPre, E., Luh, W.-M. & Spreng, R. N. (2016) Multi-echo fMRI replication sample of autobiographical memory, prospection and theory of mind reasoning tasks. *Scientific Data* **3**:160116. [aMG]
Eastwick, P. W., Finkel, E. J. & Eagly, A. H. (2011) When and why do ideal partner preferences affect the process of initiating and maintaining romantic relationships? *Journal of Personality and Social Psychology* **101**:1012–32. [AL]
Eastwick, P. W., Luchies, L. B., Finkel, E. J. & Hunt, L. L. (2014) The predictive validity of ideal partner preferences: A review and meta-analysis. *Psychological Bulletin* **140**:623–65. [AL]
Eastwick, P. W., Smith, L. K. & Ledgerwood, A. (2019) How do people translate their experiences into abstract attribute preferences? *Journal of Experimental Social Psychology* **85**:103837. [AL]
Echterhoff, G. & Higgins, E. T. (2017) Creating shared reality in interpersonal and intergroup communication: The role of epistemic processes and their interplay. *European Review of Social Psychology* **28**(1):175–226. [MR-M]
Echterhoff, G. & Higgins, E. T. (in press) Shared reality: Motivated connection and motivated cognition. In: *Social psychology: Handbook of basic principles*, 3rd Edition, ed. P. Van Lange, A. W. Kruglanski & E. T. Higgins. Guilford. [MR-M]
Echterhoff, G., Higgins, E. T. & Levine, J. M. (2009) Shared reality experiencing commonality with others' inner states about the world. *Perspectives on Psychological Science* **4**(5):496–521. Available at: http://doi.org/10.1111/j.1745-6924.2009.01161.x. [aMG, MR-M]
Edwards, G., Vetter, P., McGruer, F., Petro, LS & Muckli, L. (2017) Predictive feedback to V1 dynamically updates with sensory input. *Scientific Reports* **7**(1):16538. [LSP]
Eichenbaum, H., Yonelinas, A. P. & Ranganath, C. (2007) The medial temporal lobe and recognition memory. *Annual Review of Neuroscience* **30**:123–52. [aMG]
Eimontaite, I., Goel, V., Raymont, V., Krueger, F., Schindler, I. & Grafman, J. (2018) Differential roles of polar orbital prefrontal cortex and parietal lobes in logical reasoning with neutral and negative emotional content. *Neuropsychologia* **119**:320–29. Available at: http://doi.org/10.1016/j.neuropsychologia.2018.05.014. [aMG]
Elman, J. L. (1990) Finding structure in time. *Cognitive Science* **14**(2):179–211. [CPD]
Elman, J. L. (1993) Learning and development in neural networks: The importance of starting small. *Cognition* **48**(1):71–99. [CPD]
Elman, J. L. (2009) On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science* **33**(4):547–82. [CPD]
Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996) *Rethinking innateness: A connectionist perspective on development*. MIT Press. [CPD]
Emery, N. J. & Clayton, N. S. (2004) The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science* **306**(5703):1903–1907. [TP]
Epstein, R. & Kanwisher, N. (1998) A cortical representation of the local visual environment. *Nature* **392**(6676):598–601. Available at: http://doi.org/10.1038/33402. [aMG]
Evans, J. S. T. (2003) In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences* **7**(10):454–59. Available at: http://doi.org/10.1016/j.tics.2003.08.012. [aMG]
Faber, M. & D'Mello, S. K. (2018) How the stimulus influences mind wandering in semantically rich task contexts. *Cognitive Research: Principles and Implications* **3**(1):35. https://doi.org/10.1186/s41235-018-0129-0. [MFa]

Faber, M. & Mills, C. (2018) The critical role of the hippocampus in mind wandering. *The Journal of Neuroscience* **38**(29):6439–41. https://doi.org/10.1523/JNEUROSCI.0995-18.2018. [MFa]

Fairhall, S. L. & Caramazza, A. (2013) Brain regions that represent amodal conceptual knowledge. *The Journal of Neuroscience* **33**(25):10552–58. https://doi.org/10.1523/JNEUROSCI.0051-13.2013. [aMG, ZZ]

Farah, M. J. & McClelland, J. L. (1991) A computational model of semantic memory impairment – Modality specificity and emergent category specificity. *Journal of Experimental Psychology-General* **120**(4):339–57. doi: 10.1037/0096-3445.120.4.339 [aMG]

Faja, S., Murias, M., Beauchaine, T. P. & Dawson, G. (2013) Reward-based decision making and electrodermal responding by young children with autism spectrum disorders during a gambling task. *Autism Research* **6**(6):494–505. [aMG]

Fauconnier, G. & Turner, M. (2008) *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books. [aMG]

Fedyk, M., Kushnir, T. & Xu, F. (2019) Intuitive epistemology: Children's theory of evidence. In: *Advances in experimental philosophy of science*, ed. D. A. Wilkenfeld, R. Samuels & J. R. Beebe, pp. 122–43. Bloomsbury Academic. [MFe]

Fedyk, M. & Xu, F. (2018) The epistemology of rational constructivism. *Review of Philosophy and Psychology* **9**(2):343–62. [MFe]

Fedyk, M. & Xu, F. (2019) The metaphysics of the development of the cognitive system: Why the brain cannot replace the mind. In: *Current controversies in philosophy of cognitive science*, ed. S. Cullen & S.-J. Leslie, pp. 63–82. Routledge. [MFe]

Feld, G. B. & Born, J. (2017) Sculpting memory during sleep: Concurrent consolidation and forgetting. *Current Opinion in Neurobiology* **44**:20–27. [AV]

Feldman, H. & Friston, K. (2010) Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience* **4**:215. [LSP]

Feldman, J. I., Dunham, K., Cassidy, M., Wallace, M. T., Liu, Y. & Woynaroski, T. G. (2018) Audiovisual multisensory integration in individuals with autism spectrum disorder: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews* **95**:220–34. [MKB]

Ferro, M., Ognibene, D., Pezzulo, G. & Pirrelli, V. (2010) Reading as active sensing: A computational model of gaze planning during word recognition. *Frontiers in Neurorobotics* **4**:1. [KFr]

Festinger, L. (1962) *Cognitive dissonance*. Scientific American. [aMG]

Fiedler, K. (2014) From intrapsychic to ecological theories in social psychology: Outlines of a functional theory approach. *European Journal of Social Psychology* **44**(7):657–70. [KFi]

Fiedler, K., Bluemke, M., Freytag, P., Unkelbach, C. & Koch, S. (2008) A semiotic approach to understanding the role of communication in stereotyping. In: *Stereotype dynamics: Language-based approaches to the formation, maintenance, and transformation of stereotypes*, ed. Y. Kashima, K. Fiedler, P. & Freytag, pp. 95–116. Erlbaum. [KFi]

Fiedler, K., Jung, J., Wänke, M., Alexopoulos, T. & de Molière, L. (2015) Toward a deeper understanding of the ecological origins of distance construal. *Journal of Experimental Social Psychology* **57**:78–86. [KFi]

Field, D. J. (1994) What is the goal of sensory coding. *Neural Computation* **6**(4):559–601. Available at: http://doi.org/10.1162/neco.1994.6.4.559. [aMG]

Fields, C. & Glazebrook, J. F. (2019) A mosaic of Chu spaces and Channel Theory II: Applications to object identification and mereological complexity. *Journal of Experimental and Theoretical Artificial Intelligence* **31**:237–65. [CF]

Fields, C. & Marcianò, A. (2019) Markov blankets are general physical interaction surfaces. *Physics of Life Reviews*, in press. https://doi.org/10.1016/j.plrev.2019.08.004. [CF]

Fini, C. & Borghi, A. M. (2019) Sociality to reach objects and to catch meaning. *Frontiers in Psychology* **10**:838. [AMB]

Fitch, W. T. (2014) Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews* **11**(3):329–64. https://doi.org/10.1016/j.plrev.2014.04.005. [TP]

Fitch, W. T., Von Graevenitz, A. & Nicolas, E. (2009) Bio-aesthetics, dynamics and the aesthetic trajectory: A cognitive and cultural perspective. In: *Neuroaesthetics*, pp. 59–102. Baywood Publishing Company, Inc. [TP]

Fodor, J. A. (1975) *The language of thought, vol. 5*. Harvard University Press.

Fodor, J. A. (1998) Concepts. Where cognitive science went wrong: Clarendon Press/ Oxford University Press. [aMG]

Fodor, J. A. & Pylyshyn, Z. W. (2014) *Minds without meanings: An essay on the content of concepts*. Mit Press. [aMG]

Frege, G. (1892/1952a) On concept and object. In: *Translations from the philosophical writings of Gottlob Frege*, ed. P. T. Geach & M. Black, pp. 42–55. Basil Blackwell. (Original work published in 1892) [aMG]

Frege, G. (1892/1952b) On sense and reference. In: *Translations from the philosophical writings of Gottlob Frege*, ed. P. T. Geach & M. Black, pp. 56–78. Basil Blackwell. (Original work published in 1892) [aMG]

Friston, K. (2005) A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological Sciences* **360**(1456):815–36. [aMG, LSP, MtS]

Friston, K. (2013) Life as we know it. *Journal of the Royal Society, Interface* **10** (86):20130475. [CF, KFr]

Friston, K., Brown, H. R., Siemerkus, J. & Stephan, K. E. (2016) The dysconnection hypothesis (2016). *Schizophrenia Research* **176**(2–3):83–94. https://doi.org/10.1016/j.schres.2016.07.014. [PL]

Friston, K. & Buzsáki, G. (2016) The functional anatomy of time: What and when in the brain. *Trends in Cognitive Sciences* **20**(7):500–11. [MGV]

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P. & Pezzulo, G. (2017a) Active inference: A process theory. *Neural Computation* **29**(1):1–49. [MGV]

Friston, K., Levin, M., Sengupta, B. & Pezzulo, G. (2015) Knowing one's place: A free-energy approach to pattern regulation. *Journal of the Royal Society Interface* **12**:20141383. [CF]

Friston, K., Mattout, J. & Kilner, J. (2011) Action understanding and active inference. *Biological Cybernetics* **104**:137–60. [KFr]

Friston, K. J. (2005) A theory of cortical responses. *Philosophical Transactions of the Royal Society B-Biological Sciences* **360**(1456):815–36. Available at: http://doi.org/10.1098/rstb.2005.1622. [aMG]

Friston, K. J. (2010) The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience* **11**(2):127–38. Available at: http://doi.org/10.1038/nrn2787. [aMG]

Friston, K. J., Daunizeau, J. & Kiebel, S. J. (2009) Reinforcement learning or active inference? *PLOS ONE* **4**(7):13. Available at: http://doi.org/10.1371/journal.pone.0006421. [aMG]

Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A. & Ondobaka, S. (2017b) Active inference, curiosity and insight. *Neural Computation* **29**(10):2633–83. [AV]

Friston, K. J., Parr, T. & de Vries, B. (2017c) The graphical brain: Belief propagation and active inference. *Network Neuroscience* **1**(4):381–414. doi: 10.1162/NETN_a_00018. [KFr, rMG]

Friston, K. J., Rosch, R., Parr, T., Price, C. & Bowman, H. (2017d) Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews* **77**:388–402. doi: 10.1016/j.neubiorev.2017.04.009. [KFr]

Friston, K. J., Rosch, R., Parr, T., Price, C. & Bowman, H. (2018) Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews* **90**:486–501. [MGV]

Friston, K. J., Stephan, K. E., Montague, R. & Dolan, R. J. (2014) Computational psychiatry: The brain as a phantastic organ. *Lancet Psychiatry* **1**(2):148–58. Available at: http://doi.org/10.1016/s2215-0366(14)70275-5. [aMG]

Frith, U. (1989) *Autism: Explaining the enigma*. Blackwell. [MKB]

Frith, C. D. & Frith, U. (1999) Interacting minds – A biological basis. *Science* **286** (5445):1692–95. [aMG]

Frith, U. & de Vignemont, F. (2005) Egocentrism, allocentrism, and Asperger syndrome. *Consciousness and Cognition* **14**(4):719–38. [MKB]

Frith, U. & Happé, F. (1994) Autism: Beyond "theory of mind." *Cognition* **50**(1–3):115–32. [MKB]

Fujita, K., Trope, Y., Liberman, N. & Levin-Sagi, M. (2006) Construal levels and self-control. *Journal of Personality and Social Psychology* **90**(3):351. [aMG]

Fukukura, J., Ferguson, M. J. & Fujita, K. (2013) Psychological distance can improve decision making under information overload via gist memory. *Journal of Experimental Psychology: General* **142**(3):658–65. doi: 10.1037/a0030730. [VFR]

Fuster, J. M. (2004) Upper processing stages of the perception–action cycle. *Trends in Cognitive Sciences* **8**(4):143–45. [KFr]

Gallese, V. & Goldman, A. (1998) Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* **2**(12):493–501. Available at: http://doi.org/10.1016/s1364-6613(98)01262-5. [aMG]

Gamble, B., Moreau, D., Tippett, L. J. & Addis, D. R. (2019) Specificity of future thinking in depression: A meta-analysis. *Perspectives on Psychological Science* **14**(5):816–34. https://doi.org/10.1177/1745691619851784. [ZZ]

Gandolfo, M. & Downing, P. E. (2019) Causal evidence for expression of perceptual expectations in category-selective extrastriate regions. *Current Biology* **29**(15):2496–2500.e3. https://doi.org/10.1016/j.cub.2019.06.024. [ZZ]

Gardner, M. P., Wansink, B., Kim, J. & Park, S. B. (2014) Better moods for better eating?: How mood influences food choice. *Journal of Consumer Psychology* **24**(3):320–35. [KM]

Gaut, B. (1997) Imagination, interpretation, and film. *Philosophical Studies* **89**:331–41. [DD]

Gawronski, B. (2019) Six lessons for a cogent science of implicit bias and its criticism. *Perspectives on Psychological Science* **14**:574–95. [AL]

Gentner, D. (1983) Structure-mapping: A theoretical framework for analogy. *Cognitive Science* **7**(2): 155–70. [aMG]

Gentner, D. & Markman, A. B. (1997) Structure mapping in analogy and similarity. *American Psychologist* **52**(1):45. [aMG]

Gentner, D. & Medina, J. (1998) Similarity and the development of rules. *Cognition* **65** (2-3):263–97. Available at: http://doi.org/10.1016/s0010-0277(98)00002-x. [aMG]

George, D. & Hawkins, J. (2009) Towards a mathematical theory of cortical micro-circuits. *PLOS Computational Biology* **5**(10):e1000532. [LSP]

Gershman, S. J., Blei, D. M. & Niv, Y. (2010) Context, learning, and extinction. *Psychological review*, **117**(1):197. [aMG]

Gershman, S. J. & Daw, N. D. (2017) Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology* **68**(1):101–28. https://doi.org/10.1146/annurev-psych-122414-033625. [PD, ZZ]

Ghazanfar, A. A. & Lemus, L. (2010) Multisensory integration: Vision boosts information through suppression in auditory cortex. *Current Biology* **20**(1):R22–23. doi: 10.1016/j.cub.2009.11.046. [MP]

Ghio, M., Vaghi, M. M. S. & Tettamanti, M. (2013) Fine-grained semantic categorization across the abstract and concrete domains. *PLOS ONE* **8**(6):e67090. [AMB]

Gilbert, D. T. & Wilson, T. D. (2007) Prospection: Experiencing the future. *Science* **317**(5843): 1351–54. Available at: http://doi.org/10.1126/science.1144161. [aMG]

Gilead, M., Boccagno, C., Silverman, M., Hassin, R. R., Weber, J. & Ochsner, K. N. (2016) Self-regulation via neural simulation. *Proceedings of the National Academy of Sciences of the United States of America* **113**(36):10037–42. Available at: http://doi.org/10.1073/pnas.1600159113. [aMG]

Gilead, M., Liberman, N. & Maril, A. (2012) Construing counterfactual worlds: The role of abstraction. *European Journal of Social Psychology* **42**(3):391–97. [aMG]

Gilead, M., Liberman, N. & Maril, A. (2013) The language of future-thought: An fMRI study of embodiment and tense processing. *Neuroimage* **65**:267–79. Available at: http://doi.org/10.1016/j.neuroimage.2012.09.073. [arMG]

Gilead, M., Liberman, N. & Maril, A. (2014) From mind to matter: Neural correlates of abstract and concrete mindsets. *Social Cognitive and Affective Neuroscience* **9**(5):638–45. https://doi.org/10.1093/scan/nst031. [ZZ]

Gładziejewski, P. (2015) Predictive coding and representationalism. *Synthese* 1–24. doi: 10.1007/s11229-015-0762-9. [PL, MtS]

Goff, L. M. & Roediger, H. L. (1998) Imagination inflation for action events: Repeated imaginings lead to illusory recollections. *Memory and Cognition* **26**(1):20–33. [aMG]

Goguen, J. A. (1991) A categorical manifesto. *Mathematical Structures in Computer Science* **1**:49–67. [CF]

Goldman, A. I. (2005) Imitation, mind reading, and simulation. In: *Perspectives on imitation: Imitation, human development, and culture, Vol. 79*. [rMG]

Goldsmith, M., Koriat, A. & Weinberg-Eliezer, A. (2002) Strategic regulation of grain size in memory reporting. *Journal of Experimental Psychology-General* **131**(1):73–95. Available at: http://dx.doi.org/10.1037/0096-3445.131.173. [aMG]

Gomot, M. & Wicker, B. (2012) A challenging, unpredictable world for people with autism spectrum disorder. *International Journal of Psychophysiology* **83**(2):240–47. [MKB]

Goodman, N. (1972) Seven strictures on similarity. In: N. Goodman, *Problems and projects*, pp. 437–47. Indianapolis: Bobbs-Merrill. [aMG]

Goodman, N. D., Mansinghka, V. K., Roy, D. M., Bonawitz, K. & Tenenbaum, J. B. (2012) Church: A language for generative models. *CoRR*, abs/**1206**.3255. [PD]

Goodwin, G. P. & Johnson-Laird, P. N. (2013) The acquisition of Boolean concepts. *Trends in Cognitive Sciences* **17**(3):128–33. Available at: http://doi.org/10.1016/j.tics.2013.01.007. [aMG]

Gordon, N., Tsuchiya, N., Koenig-Robert, R. & Hohwy, J. (2019) Expectation and attention increase the integration of top-down and bottom-up signals in perception through different pathways. *PLOS Biology* **17**(4):e3000233. [LSP]

Gordon, R. (1986) Folk psychology as simulation. *Mind and Language* **1**(2):158–71. [aMG]

Gottlieb, L. J., Wong, J., de Chastelaine, M. & Rugg, M. D. (2012) Neural correlates of the encoding of multimodal contextual features. *Learning & Memory* **19**(12):605–14. [aMG]

Gow, D. W. Jr. (2001) Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language* **45**(1):133–59. [MtS]

Grafman, J., Spector, L. & Rattermann, M. J. (2005) Planning and the brain. *Cognitive Psychology of Planning* 181–98. [JG]

Grafman, J., Thompson, K., Weingartner, H., Martinez, R., Lawlor, B. A. & Sunderland, T. (1991) Script generation as an indicator of knowledge representation in patients with Alzheimer's disease. *Brain Lang* **40**(3):344–58. doi: 10.1016/0093-934x(91)90134-m. [JG]

Granito, C., Scorolli, C. & Borghi, A. M. (2015) Naming a Lego world. The role of language in the acquisition of abstract concepts. *PLOS ONE* **10**(1):e0114615. [AMB]

Gregory, R. L. (1980) Perceptions as hypothesis. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **290**(1038):181–97. Available at: http://doi.org/10.1098/rstb.1980.0090. [aMG]

Gregory, R. L. (2009) *Seeing through illusions*. Oxford University Press. [JBD]

Griffin, D. & Buehler, R. (2005) Biases and fallacies, memories and predictions: Comment on Roy, Christenfeld, and McKenzie (2005). *Psychological Bulletin* **131**(5):757–60. [KM]

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A. & Tenenbaum, J. B. (2010) Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences* **14**(8):357–64. Available at: http://doi.org/10.1016/j.tics.2010.05.004. [aMG, MGV]

Griffiths, T. L., Vul, E. & Sanborn, A. N. (2012) Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science* **21**(4):263–68. https://doi.org/10.1177/0963721412447619. [ZZ]

Grzeczkowski, L., Roinishvili, M., Chkonia, E., Brand, A., Mast, F. W., Herzog, M. H. & Shaqiri, A. (2018) Is the perception of illusions abnormal in schizophrenia? *Psychiatry Research* **270**:929–39. https://doi.org/10.1016/j.psychres.2018.10.063. [PL]

Hahn, U. & Chater, N. (1998) Understanding similarity: A joint project for psychology, case-based reasoning, and law. *Artificial Intelligence Review* **12**(5):393–427. [aMG]

Halamish, V. & Liberman, N. (2017) How much information to sample before making a decision? It's a matter of psychological distance. *Journal of Experimental Social Psychology* **71**:111–16. [aMG]

Halford, G. S., Wilson, W. H. & Phillips, S. (2010) Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences* **14**(11):497–505. Available at: http://doi.org/10.1016/j.tics.2010.08.005. [aMG]

Hamilton, A. F., Brindley, R. & Frith, U. (2009) Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* **113**(1):37–44. [MKB]

Hanson, L. K., Atance, C. M. & Paluck, S. W. (2014) Is thinking about the future related to theory of mind and executive function? Not in preschoolers. *Journal of Experimental Child Psychology* **128**:120–37. doi: 10.1016/j.jecp.2014.07.006. [BO]

Hardin, C. D. & Higgins, E. T. (1996) Shared reality: How social verification makes the subjective objective. In: *Handbook of motivation and cognition: Vol. 3. The interpersonal context*, ed. R. M. Sorrentino & E. T. Higgins, pp. 28–84. Guilford Press. [MR-M]

Harsanyi, J. C. (1967) Games with incomplete information played by "Bayesian" players, I–III Part I. The basic model. *Management Science* **14**(3):159–82. [PD]

Haslam, N. & Loughnan, S. (2014) Dehumanization and infrahumanization. In: *Annual review of psychology, vol. 65*, ed. S. T. Fiske, pp. 399–423. Annual Reviews. [aMG]

Hassabis, D., Kumaran, D. & Maguire, E. A. (2007) Using imagination to understand the neural basis of episodic memory. *Journal of Neuroscience* **27**(52):14365-374. Available at: http://doi.org/10.1523/jneurosci.4549-07.2007. [aMG]

Hauk, O., Johnsrude, I. & Pulvermüller, F. (2004) Somatotopic representation of action words in human motor and premotor cortex. *Neuron* **41**(2):301–307. [aMG]

Heider, F. (1958) *The psychology of interpersonal relations*. Wiley. [aMG]

Heit, E. & Barsalou, L. W. (1996) The instantiation principle in natural categories. *Memory (Hove, England)* **4**:413–51. [AMB]

Helm, R. K., Reyna, V. F., Franz, A. A. & Novick, R. Z. (2018) Too young to plead? Risk, rationality, and plea bargaining's innocence problem in adolescents. Psychology, *Public Policy, and Law* **24**(2):180–91. doi: 10.1037/law0000156. [VFR]

Helmholtz, H. von (1856/1961) Treatise of physiological optics: Concerning the perceptions in general. In: *Classics in psychology*, ed. T. Shipley, pp. 79–127. Philosophy Library. (Original work published in 1856) (arMG)

Henrich, J. (2015) *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press. [aMG]

Higgins, E. T. (1992) Achieving "shared reality" in the communication game: A social action that creates meaning. *Journal of Language and Social Psychology* **11**:107–31. [MR-M]

Higgins, E. T. (2019) *Shared reality: What makes us strong and tears us apart*. Oxford University Press. [MR-M]

Hill, E. L. (2004) Executive dysfunction in autism. *Trends in Cognitive Sciences* **8**(1):26–32. [MKB]

Hinton, G. & Sejnowski, T., ed. (1999) *Unsupervised learning: Foundations of neural computation*. MIT Press. [PD]

Hinton, G. E., Dayan, P., Frey, B. J. & Neal, R. M. (1995) The wake-sleep algorithm for unsupervised neural networks. *Science* **268**(5214):1158–61. Available at: http://doi.org/10.1126/science.7761831. [aMG]

Hinton, G. E., Osindero, S. & Teh, Y. W. (2006) A fast learning algorithm for deep belief nets. *Neural Computation* **18**(7):1527–54. Available at: http://doi.org/10.1162/neco.2006.18.7.1527. [aMG]

Hintzman, D. L. (1986) Schema abstraction in a multiple-trace memory model. *Psychological Review* **93**(4):411–28. Available at: http://doi.org/10.1037//0033-295x.93.4.411. [aMG]

Hirst, W. & Coman, A. (2018) Building a collective memory: The case for collective forgetting. *Current Opinion in Psychology* **23**:88–92. [MR-M]

Hirst, W. & Echterhoff, G. (2012) Remembering in conversations: The social sharing and reshaping of memories. *Annual Review of Psychology* **63**:55–79. [MR-M]

Hobeika, L., Diard-Detoeuf, C., Garcin, B., Levy, R. & Volle, E. (2016) General and specialized brain correlates for analogical reasoning: A meta-analysis of functional imaging studies. *Human Brain Mapping* **37**(5):1953–69.

Hobson, J. A. & Friston, K. J. (2012) Waking and dreaming consciousness: Neurobiological and functional considerations. *Progress in Neurobiology* **98**(1):82–98. [AV]

Hobson, J. A., Hong, C. C. H. & Friston, K. J. (2014) Virtual reality and consciousness inference in dreaming. *Frontiers in Psychology* **5**:1133. [AV]

Høffding, H. (1892) *Outlines of psychology*. Macmillan. [aMG]

Hoffman, D. D. (2018) The interface theory of perception. In: *The Stevens' handbook of experimental psychology and cognitive neuroscience, vol. II*, ed. J. Serences. Wiley (Ch. 16). [CF]

Hoffman, D. D., Singh, M. & Prakash, C. (2015) The interface theory of perception. *Psychonomic Bulletin & Review* **22**:1480–1506. [CF]

Hohwy, J. (2013) *The predictive mind*. Oxford University Press. [aMG]

Hommel, B., Musseler, J., Aschersleben, G. & Prinz, W. (2001) The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences* **24**(5):849–78. Available at: http://doi.org/10.1017/s0140525×01000103. [aMG]

Holland, P. C. (2008) Cognitive versus stimulus-response theories of learning. *Learning & Behavior* **36**(3):227–41. doi: 10.3758/lb.36.3.227. [aMG]

Hopfield, J. J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **79**(8):2554–58. Available at: http://doi.org/10.1073/pnas.79.8.2554. [aMG]

Huang, S. A., Ledgerwood, A. & Eastwick, P. W. (2020) How do ideal friend preferences and interaction context affect friendship formation? Evidence for a domain-general relationship initiation process. *Social Psychological and Personality Science* **11**:226–35. [AL]

Hubel, D. H. & Wiesel, T. N. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology* **160**:106–54. [aMG]

Hubel, D. H. & Wiesel, T. N. (1968) Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology-London* **195**(1):215–43. [aMG]

Hummel, J. E. & Holyoak, K. J. (1997) Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review* **104**(3):427. [aMG]

Humphries, C., Sabri, M., Lewis, K. & Liebenthal, E. (2014) Hierarchical organization of speech perception in human auditory cortex. *Frontiers in Neuroscience* **8**(406):1–12. [MtS]

Hurley, M. M., Dennett, D. C. & Adams, R. B., Jr. (2011) *Inside jokes: Using humor to reverse-engineer the mind*. MIT Press. [TP]

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**(7600):453–58. doi: 10.1038/nature17637. [MP, rMG]

Irish, M., Addis, D. R., Hodges, J. R. & Piguet, O. (2012) Exploring the content and quality of episodic future simulations in semantic dementia. *Neuropsychologia* **50**(14):3488–95. https://doi.org/10.1016/j.neuropsychologia.2012.09.012. [ZZ]

Irish, M. & Piguet, O. (2013) The pivotal role of semantic memory in remembering the past and imagining the future. *Frontiers in Behavioral Neuroscience* **7**:Article 27. https://doi.org/10.3389/fnbeh.2013.00027. [ZZ]

Ittelson, W. H. (1952) *The Ames demonstrations in perception*. Princeton University Press. [JBD]

Jackendoff, R. (2002) *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press. [TP]

Jakobson, R., Fant, G. M. & Halle, M. (1965) *Preliminaries to speech analysis – The distinctive features and their correlates*. The MIT Press. [MtS]

Janis, I. L. (1972) *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes*. Oxford, England: Houghton Mifflin. [aMG]

Jee, B. D. & Anggoro, F. K. (2019) Relational scaffolding enhances children's understanding of scientific models. *Psychological Science* 1–16. https://doi.org/10.1177/0956797619864601. [AD]

Jenkins, H. M. & Moore, B. R. (1973) The form of the autoshaped response with food or water reinforcers. *Journal of Experimental Analysis of Behavior* **20**:163–81. [DS]

Jensen, A. R. (1998) *The g factor: The science of mental ability*. Praeger. [AD]

Jiang, X., Long, T., Cao, W., Li, J., Dehaene, S. & Wang, L. (2018) Production of supra-regular spatial sequences by macaque monkeys. *Current Biology* **28**(12):1851–1859.e4. https://doi.org/10.1016/j.cub.2018.04.047. [TP]

Jing, H. G., Madore, K. P. & Schacter, D. L. (2017) Preparing for what might happen: An episodic specificity induction impacts the generation of alternative future events. *Cognition* **169**:118–28. https://doi.org/10.1016/j.cognition.2017.08.010. [ZZ]

Johnson, A. & Redish, A. D. (2007) Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience* **27**(45):12176–89. [aMG]

Johnson, M. (1987) *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press. [MlS]

Johnson, M. K. & Raye, C. L. (1981) Reality monitoring. *Psychological Review* **88**(1):67. [aMG]

Jones, E. E. & Davis, K. E. (1965) From acts to dispositions: The attribution process in person perception. *Advances in Experimental Social Psychology* **2**(4):219–66. [aMG]

Kahana, M. J. (1996) Associative retrieval processes in free recall. *Memory and Cognition* **24**(1):103–109. [aMG]

Kahneman, D. & Lovallo, D. (1993) Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science* **39**(1):17–31. [KM]

Kahneman, D. & Tversky, A. (1982) The simulation heuristic. In: *Judgment under uncertainty: Heuristics and biases*, ed. D. Kahneman, P. Slovic & A. Tversky, pp. 201–208. Cambridge University Press. [KM]

Kaliuzhna, M., Stein, T., Rusch, T., Sekutowicz, M., Sterzer, P. & Seymour, K. J. (2019) No evidence for abnormal priors in early vision in schizophrenia. *Schizophrenia Research* **210**:245–54. https://doi.org/10.1016/j.schres.2018.12.027. [PL]

Kandel, E. R. (2001) The molecular biology of memory storage: a dialogue between genes and synapses. *Science*, **294**(5544):1030–8. [aMG]

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I. & Kwapil, T. R. (2007) For whom the mind wanders, and when an experience-sampling study of working memory and executive control in daily life. *Psychological Science* **18**(7):614–21. [MFa]

Kanwisher, N. (2010) Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences of the United States of America* **107**(25):11163–70. Available at: http://doi.org/10.1073/pnas.1005062107. [aMG]

Kelley, H. H. (1973) Processes of causal attribution. *American Psychologist* **28**(2):107–28. Available at: http://doi.org/10.1037/h0034225. [aMG]

Kemmerer, D. (2015) Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychonomic Bulletin & Review* **22**(4):1068–75. [rMG]

Kemmerer, D., Rudrauf, D., Manzel, K. & Tranel, D. (2012) Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. *Cortex* **48**(7):826–48. doi: 10.1016/j.cortex.2010.11.001. [rMG]

Kenett, Y. N., Anaki, D. & Faust, M. (2014) Investigating the structure of semantic networks in low and high creative persons. *Frontiers in Human Neuroscience* **8**:Article 407. https://doi.org/10.3389/fnhum.2014.00407. [ZZ]

Kennedy, D. P., Redcay, E. & Courchesne, E. (2006) Failing to deactivate: Resting functional abnormalities in autism. *Proceedings of the National Academy of Sciences of the United States of America* **103**(21):8275–80. [MKB]

Keramati, M., Dezfouli, A. & Piray, P. (2011) Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLOS Computational Biology* **7**:e1002055. [PD]

Keramati, M., Smittenaar, P., Dolan, R. J. & Dayan, P. (2016) Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum. *Proceedings of the National Academy of Sciences* **113**:12868–73. [PD]

Kernberg, O. (1975/1985) *Borderline conditions and pathological narcissism*. Northvale, NJ: Jason Aronson. [aMG]

Kiefer, M. & Pulvermüller, F. (2012) Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex* **48**(7):805–25. [aMG]

Killingsworth, M. A. & Gilbert, D. T. (2010) A wandering mind is an unhappy mind. *Science* **330**(6006):932. https://doi.org/10.1126/science.1192439. [MFa]

Kim, H. (2016) Default network activation during episodic and semantic memory retrieval: A selective meta-analytic comparison. *Neuropsychologia* **80**:35–46. doi: 10.1016/j.neuropsychologia.2015.11.006 [aMG]

Kirchhoff, M., Parr, T., Palacios, E., Friston, K. & Kiverstein, J. (2018) The Markov blankets of life: Autonomy, active inference and the free energy principle. *Journal of The Royal Society Interface* **15**(138):20170792. [aMG]

Kitzbichler, M. G., Khan, S., Ganesan, S., Vangel, M. G., Herbert, M. R., Hämäläinen, M. S. & Kenet, T. (2015) Altered development and multifaceted band-specific abnormalities of resting state networks in autism. *Biological Psychiatry* **77**(9):794–804. [MKB]

Klein, S. B., Loftus, J., Trafton, J. G. & Fuhrman, R. W. (1992) Use of exemplars and abstractions in trait judgments: A model of trait knowledge about the self and others. *Journal of Personality and Social Psychology* **63**(5):739–53. https://doi.org/10.1037/0022-3514.63.5.739. [ZZ]

Klinger, E. (2013) Goal commitments and the content of thoughts and dreams: Basic principles. *Frontiers in Psychology* **4**:415. https://doi.org/10.3389/fpsyg.2013.00415. [ZZ]

Klinzing, J. G., Niethard, N. & Born, J. (2019) Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience* **22**(10):1598–1610. [AV]

Knill, D. C. & Pouget, A. (2004) The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences* **27**(12):712–19. [KFr, rMG]

Knutson, K. M., Wood, J. N. & Grafman, J. (2004) Brain activation in processing temporal sequence: An fMRI study. *NeuroImage* **23**(4):1299–307. [rMG]

Koechlin, E. & Jubault, T. (2006) Broca's area and the hierarchical organization of human behavior. *Neuron* **50**(6):963–74. doi: 10.1016/j.neuron.2006.05.017. [JG]

Koelsch, S., Rohrmeier, M., Torrecuso, R. & Jentschke, S. (2013) Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences* **110**(38):15443–15448. https://doi.org/10.1073/pnas.1300272110. [TP]

Konovalova, E. & Le Mens, G. (2020) An information sampling explanation for the in-group heterogeneity effect. *Psychological Review* **127**(1):47–73. https://doi.org/10.1037/rev0000160. [KFi]

Korn, C. W., Sharot, T., Walter, H., Heekeren, H. R. & Dolan, R. J. (2014) Depression is related to an absence of optimistically biased belief updating about future life events. *Psychological Medicine* **44**(3):579–92. https://doi.org/10.1017/S0033291713001074. [ZZ]

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M. & Del Campo, E. (2011) The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General* **140**(1):14. [MM]

Kratzer, A. & Heim, I. (1998) *Semantics in generative grammar, vol. 1185*. Blackwell Oxford. [aMG]

Kruger, T., Fiedler, K., Koch, A. S. & Alves, H. (2014) Response category width as a psychophysical manifestation of construal level and distance. *Personality and Social Psychology Bulletin* **40**(4):501–12. Available at: http://doi.org/10.1177/0146167213517009. [aMG]

Kruschke, J. K. (1992) ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review* **99**(1):22. [aMG]

Kuchling, F., Friston, K., Georgiev, G. and Levin, M. (2019) Morphogenesis as Bayesian inference: A variational approach to pattern formation and control in complex biological systems. *Physics of Life Reviews*, in press. https://doi.org/10.1016/j.plrev.2019.06.001. [CF]

Kuczaj, S. A. & Daly, M. J. (1979) The development of hypothetical reference in the speech of young children. *Journal of Child Language* **6**(03):563–79. [aMG]

Kuffler, S. W. (1953) Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology* **16**(1):37–68. [aMG]

Kwan, D., Craver, C. F., Green, L., Myerson, J., Boyer, P. & Rosenbaum, R. S. (2012) Future decision-making without episodic mental time travel. *Hippocampus* **22** (6):1215–19. Available at: http://doi.org/10.1002/hipo.20981. [aMG]

Labroo, A. A. & Patrick, V. M. (2009) Providing a moment of respite: Why a positive mood helps seeing the big picture. *Journal of Consumer Research* **35**(5):800–809. [KM]

Lacan, J. (1966) Fonction et champ de la parole et du langage en psychanalyse. In: *Écrits*, pp 237–322. Éditions du Seuil. [MKB]

Lah, S. & Miller, L. (2008) Effects of temporal lobe lesions on retrograde memory: A critical review. *Neuropsychology Review* **18**(1):24.

Lahiri, A. & Reetz, H. (2010) Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics* **38**:44–59. [MtS]

Lai, V. T. & Desai, R. H. (2016) The grounding of temporal metaphors. *Cortex* **76**:43–50. [AMB]

Lake, B. M., Ullman, T. D., Tenenbaum, J. B. & Gershman, S. J. (2017) Building machines that learn and think like people. *Behavioral and Brain Sciences* **40**:25. Available at: http://doi.org/10.1017/s0140525×16001837. [aMG]

Landauer, T. K. & Dumais, S. T. (1997) A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* **104**:211–40. [MM]

LeCun, Y., Bengio, Y. & Hinton, G. (2015) Deep learning. *Nature* **521**(7553):436–44. Available at: http://doi.org/10.1038/nature14539. [aMG]

Ledgerwood, A., Eastwick, P. W. & Smith, L. K. (2018) Toward an integrative framework for studying human evaluation: Attitudes towards objects and attributes. *Personality and Social Psychology Review* **22**:378–98. [AL]

Lee, W. S. & Atance, C. M. (2016) The effect of psychological distance on children's reasoning about future preferences. *PLOS ONE* **11**(10):e0164382. doi: 10.1371/journal.pone.0164382. [BO]

Lenci, A., Lebani, G. E. & Passaro, L. C. (2018) The emotions of abstract words: A distributional semantic analysis. *Topics in Cognitive Science* **10**(3):550–72. [MM]

Lengyel, M. and Dayan, P. (2007) Hippocampal contributions to control: The third way. In: *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 3–6, 2007*, pp. 889–96. [PD]

Lerdahl, F. & Jackendoff, R. (1983) *A generative theory of tonal music*. MIT Press. [TP]

Leslie, A. M. (1987) Pretense and representation: The origins of theory of mind. *Psychological Review* **94**(4):412–26. Available at: http://doi.org/10.1037/0033-295x.94.4.412. [aMG]

Levy, D. J. & Glimcher, P. W. (2012) The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology* **22**:1027–38. [DS]

Lewis, P. A., Knoblich, G. & Poe, G. (2018) How memory replay in sleep boosts creative problem-solving. *Trends in Cognitive Sciences* **22**(6):491–503. [AV]

Liberman, N. & Trope, Y. (1998) The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *Journal of Personality and Social Psychology* **75**(1):5. [aMG]

Liberman, N., Sagristano, M. D. & Trope, Y. (2002) The effect of temporal distance on level of mental construal. *Journal of Experimental Social Psychology* **38**(6):523–34. Available at: http://doi.org/10.1016/s0022-1031(02)00535-8. [aMG]

Liberman, N. & Trope, Y. (2008) The psychology of transcending the here and now. *Science* **322**(5905):1201–1205. [aMG]

Liberman, N. & Trope, Y. (2014) Traversing psychological distance. *Trends in Cognitive Sciences* **18**(7):364–69. Available at: http://doi.org/10.1016/j.tics.2014.03.001. [aMG, KFi]

Linville, P. W., Fischer, G. W. & Yoon, C. (1996) Perceived covariation among the features of ingroup and outgroup members: The outgroup covariation effect. *Journal of Personality and Social Psychology* **70**(3):421–36. [KFi]

Littman, M. L., Sutton, R. S. & Singh, S. P. (2001) Predictive representations of state. In: *Advances in Neural Information Processing Systems 14 [Neural Information Processing Systems: Natural and Synthetic, NIPS 2001, December 3–8, 2001*, Vancouver, British Columbia, Canada], pp. 1555–61. [PD]

Liu, N. & Bergen, B. (2016) When do language comprehenders mentally simulate locations? *Cognitive Linguistics* **27**(2):181–203. https://doi.org/10.1515/cog-2015-0123. [MlS]

Llewellyn, S. (2016) Dream to predict? REM dreaming as prospective coding. *Frontiers in Psychology* **6**:1961. [AV]

Loewenstein, G. (1996) Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes* **65**(3):272–92. [KM]

Loewenstein, G., O'Donoghue, T. & Rabin, M. (2003) Projection bias in predicting future utility. *Quarterly Journal of Economics* **118**(4):1209–48. [KM]

Lømo T (1966) Frequency potentiation of excitatory synaptic activity in the dentate area of the hippocampal formation. *Acta Physiologica Scandinavica*. **68**:128. [aMG]

Lombrozo, T. (2019) "Learning by thinking" in science and in everyday life. In: *The scientific imagination*, ed. P. Godfrey-Smith & A. Levy, pp. 230–49. Oxford University Press. [MFe]

Louwerse, M. M. (2008) Embodied relations are encoded in language. *Psychonomic Bulletin and Review* **15**(4):838–44. [CPD]

Lupyan, G. (2012) Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology* **3**:54. [CPD]

Lutz, N. D., Wolf, I., Hübner, S., Born, J. & Rauss, K. (2018) Sleep strengthens predictive sequence coding. *Journal of Neuroscience* **38**(42):8989–9000. [AV]

Mach, E. (1897) Über Gedankenexperimente. *Zeitschrift für den Physikalischen und Chemischen Unterricht* **1**:1–5. [DD]

MacKay, D. (1956) The epistemological problem for automata. In: *Automata studies*, ed. C. Shannon & J. McCarthy, pp. 235–51. Princeton University Press. [PD]

MacKay, D. J. C. (1992) Information-based objective functions for active data selection. *Neural Computation* **4**(4):590–604. doi: 10.1162/neco.1992.4.4.590. [KFr]

Macleod, C. M. (1991) Half a century of research on the stroop effect: An Integrative review. *Psychological Bulletin* **109**(2):163–203. Available at: http://doi.org/10.1037//0033-2909.109.2.163. [aMG]

Madore, K. P., Jing, H. G. & Schacter, D. L. (2019) Episodic specificity induction and scene construction: Evidence for an event construction account. *Consciousness and Cognition* **68**:1–11. https://doi.org/10.1016/j.concog.2018.12.001. [ZZ]

Madore, K. P., Szpunar, K. K., Addis, D. R. & Schacter, D. L. (2016) Episodic specificity induction impacts activity in a core brain network during construction of imagined future experiences. *Proceedings of the National Academy of Sciences of the United States of America* **113**(38):10696–701. Available at: http://doi.org/10.1073/pnas.1612278113. [aMG]

Mahler, W. (1933) Ersatzhandlungen verschiedenen realitätsgrades. *Psychologische Forschung* **18**(1):27–89. [aMG]

Majdandzic, J., Amashaufer, S., Hummer, A., Windischberger, C. & Lamm, C. (2016) The selfless mind: How prefrontal involvement in mentalizing with similar and dissimilar others shapes empathy and prosocial behavior. *Cognition* **157**:24–38. Available at: http://doi.org/10.1016/j.cognition.2016.08.003. [aMG]

Malle, B. F. & Holbrook, J. (2012) Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *Journal of Personality and Social Psychology* **102**(4):661. [aMG]

Marcus, G. (2009) *Kluge: The haphazard evolution of the human mind*. Houghton Mifflin Harcourt. [arMG]

Marcus, G. F. (2001) *The algebraic mind: Integrating connectionism and cognitive science*. MIT Press. [aMG]

Marcus, G. F., Vijayan, S., Rao, S. B. & Vishton, P. M. (1999) Rule learning by seven-month-old infants. *Science* **283**(5398):77–80. [aMG]

Markman, A. B. & Dietrich, E. (2000) In defense of representation. *Cognitive Psychology* **40**(2):138–71. [MtS]

Markman, A. B. & Stilwell, C. H. (2001) Role-governed categories. *Journal of Experimental and Theoretical Artificial Intelligence* **13**(4):329–58. [aMG]

Markus, H. (1977) Self-schemata and processing information about the self. *Journal of Personality and Social Psychology* **35**(2):63–78. https://doi.org/10.1037/0022-3514.35.2.63. [ZZ]

Markus, H. R. & Kitayama, S. (1991) Culture and the self: Implications for cognition, emotion and motivation. *Psychological Review* **98**(2):224–53. [MKB]

Marr, D. (1970) A theory for cerebral neocortex. *Proceedings of the Royal Society B: Biological Sciences* **176**:161–234. [PD]

Marr, D. (1971) Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London B* **262**:23–81. [MFa]

Marr, D. (1982) *Vision: A computational investigation into the human representation and processing of visual information*. Henry Holt. [aMG]

Martinelli, P., Sperduti, M. & Piolino, P. (2013) Neural substrates of the self-memory system: New insights from a meta-analysis. *Human Brain Mapping* **34**(7):1515–29. [aMG]

Martinho, A. & Kacelnik, A. (2016) Ducklings imprint on the relational concept of "same or different." *Science* **353**(6296):286–88. Available at: http://doi.org/10.1126/science.aaf4247. [aMG]

Masuda, T. & Nisbett, R. E. (2006) Culture and change blindness. *Cognitive Science* **30** (2):381–99. [MKB]

Mazzuca, C., Lugli, L., Benassi, M., Nicoletti, R. & Borghi, A. M. (2018) Abstract, emotional and concrete concepts and the activation of mouth-hand effectors. *PeerJ* **6**: e5987. [AMB]

McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. (1995) Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review* **102**(3):419. [aMG]

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S. & Smith, L. B. (2010) Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences* **14**(8):348–56. Available at: http://doi.org/10.1016/j.tics.2010.06.002. [aMG]
McClelland, J. L., Rumelhart, D. E. & PDP Research Group. (1986) *Parallel distributed processing*. MIT Press. [aMG]
McCormick, C., Rosenthal, C. R., Miller, T. D. & Maguire, E. A. (2018) Mind-wandering in people with hippocampal damage. *The Journal of Neuroscience* **38**(11):2745–54. https://doi.org/10.1523/JNEUROSCI.1812-17.2018. [MFa]
McNamara, J. M. & Houston, A. I. (1986) The common currency for behavioral decisions. *The American Naturalist* **127**:358–78. [DS]
Mead, G. H. (1934) *Mind, self, and society: From the standpoint of a social behaviorist: Works of George Herbert Mead, vol. 1*. 111. Chicago: University of Chicago Press. [aMG]
Medin, D. L. & Schaffer, M. M. (1978) Context theory of classification learning. *Psychological review* **85**(3):207. [aMG]
Medin, D. L. (1989) Concepts and conceptual structure. *American Psychologist* **44** (12):1469–81. [MtS]
Medin, D. L., Goldstone, R. L. & Gentner, D. (1993) Respects for similarity. *Psychological Review* **100**(2):254. [aMG]
Mednick, S. (1962) The associative basis of the creative process. *Psychological Review* **69** (3):220–32. https://doi.org/10.1037/h0048850. [ZZ]
Melloni, L., Buffalo, E. A. & Dehaene, S. (2019) Computation and its neural implementation in human cognition. In: Strüngmann Forum Reports: Vol. 27. *The Neocortex*, pp. 323–46. MIT Press. [MGV]
Meyer, M. L., Hershfield, H. E., Waytz, A. G., Mildner, J. N. & Tamir, D. I. (2019) Creative expertise is associated with transcending the here and now. *Journal of Personality and Social Psychology* **116**(4):483–94. https://doi.org/10.1037/pspa0000148. [rMG, ZZ]
Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013) Efficient estimation of word representations in vector space. *arXiv preprint arXiv*:**1301**.3781. [rMG]
Mildner, J. N. & Tamir, D. I. (2019) Spontaneous thought as an unconstrained memory process. *Trends in Neurosciences* **42**(11):8. [ZZ]
Miller, G. A., Galanter, E. & Pribram, K. H. (1986) *Plans and the structure of behavior*. Adams-Bannister-Cox. [JG]
Milligan, K., Astington, J. W. & Dack, L. A. (2007) Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development* **78**(2):622–46. [aMG]
Mills, B. A., Reyna, V. F. & Estrada, S. (2008) Explaining contradictory relations between risk perception and risk taking. *Psychological Science* **19**(5):429–34. doi: 10.1111/j.1467-9280.2008.02104.x. [VFR]
Mills, C., Herrera-Bennett, A., Faber, M. & Christoff, K. (2018) Why the mind wanders: How spontaneous thought's default variability may support episodic efficiency and semantic optimization. In: *The Oxford handbook of spontaneous thought: Mind-wandering, creativity, and dreaming*, ed. K. Christoff & K. C. R. Fox, pp. 11–22. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190464745.013.42. [MFa]
Mills, C., Raffaelli, Q., Irving, Z. C., Stan, D. & Christoff, K. (2017) Is an off-task mind a freely-moving mind? Examining the relationship between different dimensions of thought. *Consciousness and Cognition* **58**:20–33. https://doi.org/10.1016/j.concog.2017.10.003. [MFa]
Moessinger, P. & Poulin-Dubois, D. (1981) Piaget on abstraction. *Human Development* **24**(5):347–53. doi: 10.1159/000272712. [JTB]
Montefinese, M. (2019) Semantic representation of abstract and concrete words: A minireview of neural evidence. *Journal of neurophysiology* **121**(5):1585–87. [aMG, MM]
Montefinese, M., Ambrosini, E., Fairfield, B. & Mammarella, N. (2013) The "subjective" pupil old/new effect: Is the truth plain to see? *International Journal of Psychophysiology* **89**(1):48–56. [MM]
Mormann, F., Kornblith, S., Quiroga, R. Q., Kraskov, A., Cerf, M., Fried, I. & Koch, C. (2008) Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *Journal of Neuroscience* **28**(36):8865–72. Available at: http://doi.org/10.1523/jneurosci.1640-08.2008. [aMG]
Möttönen, R., Dutton, R. & Watkins, K. E. (2013) Auditory-motor processing of speech sounds. *Cerebral Cortex* **23**(5):1190–97. [MtS]
Möttönen, R. & Watkins, K. E. (2009) Motor representations of articulators contribute to categorical perception of speech sounds. *The Journal of Neuroscience* **29**(31):9819–25. [MtS]
Moulton, S. T. & Kosslyn, S. M. (2009) Imagining predictions: Mental imagery as mental emulation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**(1521):1273–80. [aMG]
Mrkva, K., Travers, M. & Van Boven, L. (2018) Simulational fluency reduces feelings of psychological distance. *Journal of Experimental Psychology: General* **147**(3):354–76. [KM]
Mrkva, K. & Van Boven, L. (2017) Attentional accounting: Voluntary spatial attention increases budget category prioritization. *Journal of Experimental Psychology: General* **146**(9):1296–306. [KM]
Murphy, G. L. & Medin, D. L. (1985) The role of theories in conceptual coherence. *Psychological Review* **92**(3):289. [arMG]
Murphy, G. L. & Wisniewski, E. J. (1989) Categorizing objects in isolation and in scenes: What a superordinate is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15**(4):572. [AMB]
Narayanan, S. (1999) Reasoning about actions in narrative understanding. In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'1999)*, pp. 350–57. Morgan Kaufmann Press. [MlS]
Neisser, U. (1967) *Cognitive psychology*. Appleton-Century-Crofts. [PD]
Newell, A. & Simon, H. A. (1972) *Human problem solving*. Prentice Hall. [aMG]
Ng, A. Y. & Jordan, M. I. (2002) On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In: *Advances in neural information processing systems*, ed. T. G. Dietterich, S. Becker & Z. Ghahramani, pp. 841–48. [arMG]
Nichelli, P., Grafman, J., Pietrini, P., Alway, D., Carton, J. C. & Miletich, R. (1994) Brain activity in chess playing. *Nature* **369**(6477):191. doi: 10.1038/369191a0. [JG]
Nichelli, P., Grafman, J., Pietrini, P., Clark, K., Lee, K. Y. & Miletich, R. (1995) Where the brain appreciates the moral of a story. *NeuroReport* **6**(17):2309–2313. doi: 10.1097/00001756-199511270-00010. [JG]
Nichols, S. & Stich, S. (2000) A cognitive theory of pretense. *Cognition* **74**(2):115–47. [aMG]
Nisbett, R. E. & Masuda, T. (2003) Culture and point of view. *Proceedings of the National Academy of Sciences of the United States of America* **100**(19):11163–70. [MKB]
Niv, Y., Joel, D. & Dayan, P. (2006) A normative perspective on motivation. *Trends in Cognitive Sciences* **10**(8):375–81. [aMG]
Noordhof, P. (2002) Imagining objects and imagining experiences. *Mind and Language* **17**:426–55. [DD]
Norenzayan, A. (2013) *Big gods: How religion transformed cooperation and conflict*. Princeton University Press. [aMG]
Northoff, G. (2018) How does the brain's spontaneous activity generate our thoughts? In: *The Oxford handbook of spontaneous thought*, ed. K. Christoff & K. C. R. Fox, pp 55–70. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190464745.013.9. [ZZ]
Obleser, J. & Eisner, F. (2009) Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences* **13**(1):14–19. [MtS]
Okasha, S. (2013) The evolution of Bayesian updating. *Philosophy of Science* **80** (5):745–57. [DS]
O'Reilly, R. C. & McClelland, J. L. (1994) Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus* **4**(6):661–82. Available at: http://doi.org/10.1002/hipo.450040605. [aMG]
Ostrolenk, A., Bao, V. A., Mottron, L., Collignon, O. & Bertone, A. (2019) Reduced multisensory facilitation in adolescents and adults on the autism spectrum. *Scientific Reports* **9**(1):11965. [MKB]
Oudiette, D. & Paller, K. A. (2013) Upgrading the sleeping brain with targeted memory reactivation. *Trends in Cognitive Sciences* **17**(3):142–49. [AV]
Packer, D. J. & Cunningham, W. A. (2009) Neural correlates of reflection on goal states: The role of regulatory focus and temporal distance. *Social Neuroscience* **4**(5):412–25. [aMG]
Paivio, A. (1971) *Imagery and verbal processes*. Holt, Rinehart, & Winston. [MM]
Papageorgiou, E., Christou, C., Spanoudis, G. & Demetriou, A. (2016) Augmenting intelligence: Developmental limits to learning-based cognitive change. *Intelligence* **56**:16–27. [AD]
Parducci, A. (1965) Category judgment: A range-frequency model. *Psychological Review* **72**(6):407–418. [KFi]
Parr, T. & Friston, K. J. (2019) Generalised free energy and active inference. *Biological Cybernetics*. doi: 10.1007/s00422-019-00805-w. [JG]
Peacocke, C. (1985) Imagination, experience and possibility. In: *Essays on Berkeley*, ed. J. Forster & H. Robinson, pp. 19–35. Oxford University Press. [DD]
Pearce, J. M. (1987) A model for stimulus-generalization in Pavlovian conditioning. *Psychological Review* **94**(1):61–73. Available at: http://doi.org/10.1037//0033-295x.94.1.61. [aMG]
Pearson, A., Marsh, L., Hamilton, A. F. & Ropar, D. (2014) Spatial transformations of bodies and objects in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders* **44**(9):2277–89. [MKB]
Pearson, J., Naselaris, T., Holmes, E. A. & Kosslyn, S. M. (2015) Mental imagery: Functional mechanisms and clinical applications. *Trends in Cognitive Sciences* **19** (10):590–602. Available at: http://doi.org/10.1016/j.tics.2015.08.003. [aMG]
Peetz, J. & Buehler, R. (2009) Is there a budget fallacy? The role of savings goals in the prediction of personal spending. *Personality and Social Psychology Bulletin* **35** (12):1579–91. [KM]
Peirce, C. S. (1931) *Collected papers of Charles Sanders Peirce*. Harvard University Press. [aMG]
Penn, D. C., Holyoak, K. J. & Povinelli, D. J. (2008) Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences* **31** (2):109–30. [aMG, TP]
Peters, J. & Büchel, C. (2010) Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. *Neuron* **66** (1):138–48. [arMG]
Pezzulo, G., Kemere, C. & van der Meer, M. A. A. (2017) Internally generated hippocampal sequences as a vantage point to probe future-oriented cognition. *Annals of the*

New York Academy of Sciences **1396**(1):144–65. Available at: http://doi.org/10.1111/nyas.13329. [aMG, AV]

Pezzulo, G., Rigoli, F. & Chersi, F. (2013) The mixed instrumental controller: Using value of information to combine habitual choice and mental simulation. *Frontiers in Psychology* **4**:92. [PD]

Piaget, J. (1954) *The construction of reality in the child*. Basic Books. [aMG]

Piaget, J. (1976/1979) *Behaviour and evolution* (D. Nicholson-Smith, trans.). Routledge & Kegan Paul. (Original work published in 1976) [JTB]

Piaget, J. (1977/2001) *Studies in reflecting abstraction* (R. L. Campbell, trans.). Taylor & Francis. (Original work published in 1977) [JTB]

Piaget, J. (1980a) *Les formes élémentaires de la dialectique*. Gallimard. [JTB]

Piaget, J. (1980b) *Recherches sur les correspondances*. Presses Universitaires de France. [JTB]

Piaget, J. (1981/1987) *Possibility and necessity: The role of possibility in cognitive development* (H. Feider, trans. Vol. 1). University of Minnesota Press. (Original work published in 1981) [JTB]

Piaget, J. (1983/1987) *Possibility and necessity: The role of necessity in cognitive development* (H. Feider, trans. Vol. 2). University of Minnesota Press. (Original work published in 1983) [JTB]

Piaget, J. (2001) *Studies in reflecting abstraction*. Psychology Press. [AD]

Piaget, J. & Garcia, R. (1983/1989) *Psychogenesis and the history of science* (H. Feider, trans.). Columbia University Press. (Original work published in 1983) [JTB]

Piaget, J. & Garcia, R. (1987/1991) *Toward a logic of meanings* (P. M. Davidson & J. Easley, eds.; D. De Caprona & P. M. Davidson, trans.). Lawrence Erlbaum. (Original work published in 1987) [JTB]

Piaget, J. & Henriques, G. (1978) *Recherches sur la généralisation*. Presses universitaires de France. [JTB]

Piaget, J., Henriques, G. & Ascher, E. (1990/1992) *Morphisms and categories: Comparing and transforming* (T. Brown, trans.). Lawrence Erlbaum. (Original work published in 1990) [JTB]

Pick-Alony, R., Liberman, N. & Trope, Y. (2014) High level of construal and psychological distance reduce melioration. *Journal of Behavioral Decision Making* **27**(4):291–300. [aMG]

Pickering, M. J. & Garrod, S. (2013) An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* **36**(4):329–47. [CPD]

Pinker, S. (2007) *The stuff of thought: Language as a window into human nature*. Penguin. [aMG]

Pleskac, T. J. & Hertwig, R. (2014) Ecologically rational choice and the structure of the environment. *Journal of Experimental Psychology: General* **143**(5):2000–2019. [KFi]

Poeppel, D. (2012) The maps problem and the mapping problem: Two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology* **29**(1–2):34–55. [MGV]

Pollock, L. (2018) Statistical and methodological problems with concreteness and other semantic variables: A list memory experiment case study. *Behavior Research Methods* **50**(3):1198–216. [AMB]

Postle, N., McMahon, K. L., Ashton, R., Meredith, M. & de Zubicaray, G. I. (2008) Action word meaning representations in cytoarchitectonically defined primary and premotor cortices. *NeuroImage* **43**(3):634–44. [rMG]

Pothos, E. M. & Chater, N. (2002) A simplicity principle in unsupervised human categorization. *Cognitive Science* **26**(3):303–43. [aMG]

Powers, A. R., Mathys, C. & Corlett, P. R. (2017) Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science* **357**(6351):596–600. https://doi.org/10.1126/science.aan3458. [aMG, PL]

Prencipe, A. & Zelazo, P. D. (2005) Development of affective decision making for self and other: Evidence for the integration of first-and third-person perspectives. *Psychological Science* **16**:501–505. doi: 10.1111/j.0956-7976.2005.01564.x. [BO]

Prinz, J. (2012a) *Regaining composure: A defense of prototype compositionality*. *The Oxford Handbook of Compositionality* 437–53. [aMG]

Prinz, J. J. (2012b) *Beyond human nature. How culture and experience shape our lives*. Penguin; Norton. [AMB]

Proctor, R. W. & Cho, Y. S. (2006) Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin* **132**(3):416–42. [KFi]

Pulvermüller, F. (2018) Neural reuse of action perception circuits for language, concepts and communication. *Progress in Neurobiology* **160**:1–44. https://doi.org/10.1016/j.pneurobio.2017.07.001. [MlS]

Pylkkänen, L., Brennan, J. & Bemis, D. K. (2011) Grounding the cognitive neuroscience of semantics in linguistic theory. *Language and Cognitive Processes* **26**(9):1317–37. [aMG]

Pylyshyn, Z. W. (1978) When is attribution of beliefs justified? [P&W]. *Behavioral and Brain Sciences* **1**(4):592–93. [aMG]

Quiroga, R. Q., Kraskov, A., Koch, C. & Fried, I. (2009) Explicit encoding of multimodal percepts by single neurons in the human brain. *Current Biology* **19**(15):1308–13. Available at: http://doi.org/10.1016/j.cub.2009.06.060. [aMG]

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C. & Fried, I. (2005) Invariant visual representation by single neurons in the human brain. *Nature* **435**(7045):1102–1107. Available at: http://doi.org/10.1038/nature03687. [aMG]

Radvansky, G. A. & Zacks, J. M. (2017) Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences* **17**:133–40. doi: 10.1016/j.cobeha.2017.08.006. [JG]

Raichle, M. E. (2015) The brain's default mode network. *Annual Review of Neuroscience* **38**:433–47. [MKB]

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A. & Shulman, G. L. (2001) A default mode of brain function. *Proceedings of the National Academy of Sciences* **98**(2):676–82. [aMG]

Raju, P. C. (in preparation) Levels of kernel abstraction: A proposed framework for the neurobiological basis of cognition. *Computational Psychiatry*. [AD]

Ramsey, W. M. (2010) *Representation reconsidered*. Cambridge University Press. [MtS]

Rao, R. P. N. & 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. Available at: http://doi.org/10.1038/4580. [aMG, KFr, PD]

Raposo, A., Mendes, M. & Marques, J. F. (2012) The hierarchical organization of semantic memory: Executive function in the processing of superordinate concepts. *Neuroimage* **59**(2):1870–78. Available at: http://doi.org/10.1016/j.neuroimage.2011.08.072. [aMG]

Rasch, B. & Born, J. (2013) About sleep's role in memory. *Physiological Reviews* **93**(2):681–766. [AV]

Rasmussen, J. (1985) The role of hierarchical knowledge representation in decision making and system management. *IEEE Transactions on Systems, Man, and Cybernetics* **15**(2):234–43. [DAB]

Ratcliff, M. J. & Burman, J. T. (2017) The mobile frontiers of Piaget's psychology: From academic tourism to interdisciplinary collaboration/Las fronteras móviles de la psicología de Piaget. Del turismo académico a la colaboración interdisciplinaria. [Spanish translation by Julia Fernández Treviño.]. *Estudios de Psicología: Studies in Psychology* **38**(1):4–36. doi: 10.1080/02109395.2016.1268393. [JTB]

Rattermann, M. J., Spector, L., Grafman, J., Levin, H. & Harward, H. (2001) Partial and total-order planning: evidence from normal and prefrontally damaged populations. *Cognitive Science* **25**(6):941–75, doi: 10.1016/S0364-0213(01)00060-X. [JG]

Recchia, G. & Louwerse, M. M. (2015) Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance. *The Quarterly Journal of Experimental Psychology* **68**:1584–98. [MM]

Redish, A. D. (2013) *The mind within the brain: How we make decisions and how those decisions go wrong*. Oxford University Press. [aMG]

Redish, A. D. (2016) Vicarious trial and error. *Nature Reviews Neuroscience* **17**(3):147–59. Available at: http://doi.org/10.1038/nrn.2015.30. [aMG]

Redish, A. D., Jensen, S., Johnson, A. & Kurth-Nelson, Z. (2007) Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling. *Psychological Review* **114**(3):784–805. Available at: http://doi.org/10.1037/0033-295x.114.3.784. [aMG]

Reed, S. K. (2016) A taxonomic analysis of abstraction. *Perspectives on Psychological Science* **11**:817–37. [AD]

Reeves, L. M. & Weisberg, R. W. (1994) The role of content and abstract information in analogical transfer. *Psychological Bulletin* **115**(3):381–400. Available at: http://doi.org/10.1037//0033-2909.115.3.381. [aMG]

Renoult, L., Davidson, P. S. R., Palombo, D. J., Moscovitch, M. & Levine, B. (2012) Personal semantics: At the of semantic and episodic memory. *Trends in Cognitive Sciences* **16**(11):550–58. https://doi.org/10.1016/j.tics.2012.09.003. [ZZ]

Renoult, L., Kopp, L., Davidson, P. S. R., Taler, V. & Atance, C. M. (2016) You'll change more than I will: Adults' predictions about their own and others' future preferences. *The Quarterly Journal of Experimental Psychology* **69**:299–309. doi: 10.1080/17470218.2015.1046463. [BO]

Reyna, V. F. (2012) A new intuitionism: Meaning, memory, and development in fuzzy-trace theory. *Judgment and Decision Making* **7**(3):332–59. [VFR]

Reyna, V. F. & Brainerd, C. J. (2011) Dual processes in decision making and developmental neuroscience: A fuzzy-trace model. *Developmental Review* **31**:180–206. [VFR]

Reyna, V. F., Estrada, S. M., DeMarinis, J. A., Myers, R. M., Stanisz, J. M. & Mills, B. A. (2011) Neurobiological and memory models of risky decision making in adolescents versus young adults. Journal of Experimental Psychology: Learning, *Memory, and Cognition* **37**(5):1125–42. [VFR]

Reyna, V. F., Helm, R. K., Weldon, R. B., Shah, P. D., Turpin, A. G. & Govindgari, S. (2018) Brain activation covaries with reported criminal behaviors when making risky choices: A fuzzy-trace theory approach. *Journal of Experimental Psychology: General* **147**(7):1094–1109. [VFR]

Reyna, V. F. & Panagiotopoulos, C. (in press) Morals, money, and risk taking from childhood to adulthood: The neurodevelopmental framework of fuzzy-trace theory. In: *The social brain – a developmental perspective*, ed. J. Decety. The MIT Press. [VFR]

Ring, M., Gaigg, S. B., Altgassen, M., Barr, P. & Bowler, D. M. (2018) Allocentric versus egocentric spatial memory in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders* **48**(6):2101–11. [MKB]

Rodriguez, M. L., Mischel, W. & Shoda, Y. (1989) Cognitive person variables in the delay of gratification of older children at risk. *Journal of Personality and Social Psychology* **57** (2):358. [aMG]

Roelofs, A. (1992) A spreading-activation theory of lemma retrieval in speaking. *Cognition* **42** (1–3): 107–42. Available at: http://doi.org/10.1016/0010-0277(92)90041-f. [aMG, MR-M]

Rogers, T. T. & McClelland, J. L. (2004) *Semantic cognition: A parallel distributed processing approach*. MIT Press. [rMG]

Rogoff, B. (1990) *Apprenticeship in thinking: Cognitive development in social context*. Oxford University Press. [MFe]

Rolls, E. T. (2013) *Emotion and decision-making explained*. Oxford University Press. [DS]

Rosch, E. (1973) Natural categories. *Cognitive Psychology* **4**(3):328–50. [MtS]

Rosch, E. & Mervis, C. B. (1975) Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* **7**(4):573–605. [aMG]

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyesbraem, P. (1976) Basic objects in natural categories. *Cognitive Psychology* **8**(3):382–439. Available at: http://doi.org/10.1016/0010-0285(76)90013-x. [aMG]

Rosen, D. A. (1975) An argument for the logical notion of a memory trace. *Philosophy of Science* **42**(1):1–10. [MtS]

Rosen, V. M., Caplan, L., Sheesley, L., Rodriguez, R. & Grafman, J. (2003) An examination of daily activities and their scripts across the adult lifespan. *Behavior Research Methods, Instruments, & Computers* **35**(1):32–48. [JG]

Rossignac-Milon, M. & Higgins, E. T. (2018a) Beyond intrapersonal cognitive consistency: Shared reality and the interpersonal motivation for truth. *Psychological Inquiry* **29**(2):86–93. [MR-M]

Rossignac-Milon, M. & Higgins, E. T. (2018b) Epistemic companions: The development of shared reality in close relationships. *Current Opinion in Psychology* **23**:66–71. [MR-M]

Rotaru, A. S., Vigliocco, G. & Frank, S. L. (2018) Modeling the structure and dynamics of semantic processing. *Cognitive Science* **42**(8):2890–917. [MM]

Ruffman, T., Slade, L. & Crowe, E. (2002) The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development* **73** (3):734–51. Available at: http://doi.org/10.1111/1467-8624.00435. [aMG]

Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J. & Daw, N. D. (2017) Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLOS Computational Biology* **13**(9):e1005768. doi: 10.1371/journal.pcbi.1005768. [KFr]

Rutt, J. L., O'Brien, B. M. & Atance, C. M. (2019) *Spatial distance and reasoning about the future*. Poster presented at the Biennial Meeting of the Society for Research in Child Development, Baltimore, MD. [BO]

Ryali, S., Supekar, K., Chen, T., Kochalka, J., Cai, W., Nicholas, J., Padmanabhan, A. & Menon, V. (2016) *PLOS Computational Biology* **12**(12):e1005138. [MKB]

Saitta, L. & Zucker, J.-D. (2013) *Abstraction in artificial intelligence and complex systems*. Springer. [AD]

Sakreida, K., Scorolli, C., Menz, M. M., Heim, S., Borghi, A. M. & Binkofski, F. (2013) Are abstract action words embodied? An fMRI investigation at the interface between language and motor cognition. *Frontiers in Human Neuroscience* **7**:125. [AMB]

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B. L., Witter, M. P., Moser, M. B. & Moser, E. I. (2006) Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science*: **312**(5774):758–62. [aMG]

Savage-Rumbaugh, E. S., Rumbaugh, D. M. & Boysen, S. (1978) Symbolic communication between two chimpanzees (Pan troglodytes). *Science* **201**(4356):641–44. [aMG]

Sayette, M. A., Loewenstein, G., Griffin, K. M. & Black, J. J. (2008) Exploring the cold-to-hot empathy gap in smokers. *Psychological Science* **19**(9):926–32. [KM]

Schacter, D. L., Addis, D. R. & Buckner, R. L. (2007) Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience* **8**(9):657–61. Available at: http://doi.org/10.1038/nrn2213. [aMG]

Schank, R. C. & Abelson, R. P. (1975) *Scripts, plans, and knowledge*. Yale University. [aMG]

Schank, R. C. & Abelson, R. P. (1977) *Scripts, plans, goals, and understanding: an inquiry into human knowledge structures*. L. Erlbaum Associates; distributed by the Halsted Press Division of John Wiley and Sons. [JG]

Scharinger, M., Monahan, P. J. & Idsardi, W. J. (2016) Linguistic category structure influences early auditory processing: Converging evidence from mismatch responses and cortical oscillations. *NeuroImage* **128**:293–301. [MtS]

Schenker, H. (1935) *Der Freie Satz. Neue musikalische Theorien und Phantasien*. Universal Edition. [TP]

Schmack, K., Schnack, A., Priller, J. & Sterzer, P. (2015) Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. *Schizophrenia Research: Cognition* **2**(2):72–77. doi: 10.1016/j.scog.2015.03.005 [PL]

Schmidhuber, J. (2006) Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connection Science* **18**(2):173–87. doi: 10.1080/09540090600768658. [KFr]

Schultz, W., Dayan, P. & Montague, P. R. (1997) A neural substrate of prediction and reward. *Science* **275**(5306):1593–99. Available at: http://doi.org/10.1126/science.275.5306.1593. [aMG]

Schwartenbeck, P., FitzGerald, T., Dolan, R. J. & Friston, K. J. (2013) Exploration, novelty, surprise, and free energy minimization. *Frontiers in Psychology* **4**:5. Available at: http://doi.org/10.3389/fpsyg.2013.00710. [aMG]

Schwarz, N. (2001) Feelings as information: Implications for affective influences on information processing. In: *Theories of mood and cognition: A user's guidebook*, ed. L. Martin & G. Clore, pp. 159–76. Lawrence Erlbaum Associates Publishers. [KM]

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H. & Simons, A. (1991) Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology* **61**(2):195–202. [KM]

Schwarz, N. & Clore, G. L. (1996) Feelings and phenomenal experiences. *Social Psychology: Handbook of Basic Principles* **2**:385–407. [KM]

Schwiedrzik, C. M. & Freiwald, W. A. (2017) High-level prediction signals in a low-level area of the macaque face-processing hierarchy. *Neuron* **96**(1):89–97. [MGV]

Searle, J. R. (1979) What is an intentional state? *Mind* **88**(349):74–92. [aMG]

Searle, J. R. (1983) *Intentionality: An essay in the philosophy of mind*. Cambridge University Press. [aMG]

Searle, J. R. (2010) *Making the social world: The structure of human civilization*. Oxford University Press. [aMG]

Segall, M. H., Campbell, D. T. & Herskovits, M. J. (1966) *The influence of culture on visual perception*. Bobbs-Merrill, pp. 174–84. Oxford University Press. [JBD]

Semin, G. R. & Fiedler, K. (1991) The linguistic category model, its bases, applications and range. *European Review of Social Psychology* **2**(1):1–30. [aMG]

Seo, H. S., Arshamian, A., Schemmer, K., Scheer, I., Sander, T., Ritter, G. & Hummel, T. (2010) Cross-modal integration between odors and abstract symbols. *Neuroscience Letters* **478**(3):175–78. doi: 10.1016/j.neulet.2010.05.011. [MP]

Seth, A. K. (2014) A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience* **5**(2):97–118. doi: 10.1080/17588928.2013.877880. [KFr]

Seth, A. K. & Tsakiris, M. (2018) Being a beast machine: The somatic basis of selfhood. *Trends in Cognitive Sciences* **22**:969–81. [CF]

Shallice, T. & Cooper, R. P. (2013) Is there a semantic system for abstract words? *Frontiers in Human Neuroscience* **7**:175. [AMB]

Shannon, B. J., Raichle, M. E., Snyder, A. Z., Fair, D. A., Mills, K. L., Zhang, D., Bache, K., Calhoun, V. D., Nigg, J. T., Nagel, B. J., Stevens, A. A. & Kiehl, K. A. (2011) Premotor functional connectivity predicts impulsivity in juvenile offenders. *Proceedings of the National Academy of Sciences of the United States of America* **108**(27):11241–45. [MKB]

Shapira, O., Liberman, N., Trope, N. & Rim, S. (2012) Levels of mental construal. *Sage Handbook of Social Cognition*, 229–50. [aMG]

Shea, N. (2018) Metacognition and abstract concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**(1752):20170133. [AMB]

Shepard, R. N. (1987) Toward a universal law of generalization for psychological science. *Science* **237**(4820):1317–23. [aMG]

Sherman, S. J., Cialdini, R. B., Schwartzman, D. F. & Reynolds, K. D. (1985) Imagining can heighten or lower the perceived likelihood of contracting a disease: The mediating effect of ease of imagery. *Personality and Social Psychology Bulletin* **11**(1):118–27. [KM]

Shillcock, R. (2014) The concrete universal and cognitive science. *Axiomathes* **24** (1):63–80. [RS]

Shillcock, R., Thomas, J. & Bailes, R. (2019) Mirror neurons, prediction and hemispheric coordination: The prioritizing of intersubjectivity over "intrasubjectivity". *Axiomathes* **29**(2):139–53. [RS]

Siegel, J. Z., Mathys, C., Rutledge, R. B. & Crockett, M. J. (2018) Beliefs about bad people are volatile. *Nature Human Behaviour* **2**(10):750–56. https://doi.org/10.1038/s41562-018-0425-1. [ZZ]

Simons, J. S., Garrison, J. R. & Johnson, M. K. (2017) Brain mechanisms of reality monitoring. *Trends in Cognitive Sciences* **21**:462–73. [CF]

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A. & Hasson, U. (2016) Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature Communications* **7**:12141. https://doi.org/10.1038/ncomms12141. [ZZ]

Simpson, J. R. Jr., Drevets, W. C., Snyder, A. Z., Gusnard, D. A. & Raichle, M. E. (2001a) Emotion-induced changes in human medial prefrontal cortex: II. During anticipatory anxiety. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 688–93. [MKB]

Simpson, J. R. Jr., Snyder, A. Z., Gusnard, D. A. & Raichle, M. E. (2001b) Emotion-induced changes in human medial prefrontal cortex: I. During cognitive task performance. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 683–87. [MKB]

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., Diamond, S. P. & Held, R. M. (2014) Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences of the United States of America* **111**(42):15220–25. [MKB]

Sirigu, A., Zalla, T., Pillon, B., Grafman, J., Agid, Y. & Dubois, B. (1996) Encoding of sequence and boundaries of scripts following prefrontal lesions. *Cortex* **32**(2):297–310. doi: 10.1016/s0010-9452(96)80052-9. [JG]

Skinner, B. F. (1938) *The behavior of organisms: An experimental analysis*. New York: Appleton-Century-Crofts. [aMG]
Sloutsky, V. M. (2003) The role of similarity in the development of categorization. *Trends in Cognitive Sciences* **7**(6):246–51. https://doi.org/10.1016/S1364-6613(03)00109-8. [MFa]
Smith, E., Zhang, S. & Bennetto, L. (2017) Temporal synchrony and audiovisual integration of speech and object stimuli in autism. *Research in Autism Spectrum Disorders* **39**:11–19. [MKB]
Smith, E. R. (1998) *Mental representation and memory*. In: D. T. Gilbert, S. T. Fiske, G. Lindzey (eds). *The handbook of social psychology*, vol. 2, pp. 391–445. Oxford University Press. [aMG]
Smith, E. R. & Semin, G. R. (2007) Situated social cognition. *Current Directions in Psychological Science* **16**(3):132–35. [aMG]
Snefjella, B. & Kuperman, V. (2015) Concreteness and psychological distance in natural language use. *Psychological Science* **26**(9):1449–60. Available at: http://doi.org/10.1177/0956797615591771. [aMG]
Snyder, J. S., Schwiedrzik, C. M., Vitela, A. D. & Melloni, L. (2015) How previous experience shapes perception in different sensory modalities. *Frontiers in Human Neuroscience* **9**:594. [MGV]
Spearman, C. (1904) "General intelligence," Objectively determined and measured. *The American Journal of Psychology* **15**:201–92. [AD]
Spratling, M. W. (2017) A review of predictive coding algorithms. *Brain and Cognition* **112**:92–97. [LSP]
Spreng, R. N., Mar, R. A. & Kim, A. S. (2009) The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. *Journal of Cognitive Neuroscience* **21**(3):489–510. [aMG]
Spreng, R. N. & Grady, C. L. (2010) Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. *Journal of Cognitive Neuroscience* **22**(6):1112–23. [aMG]
Spreng, R. N. & Turner, G. R. (2019) The shifting architecture of cognition and brain function in older adulthood. *Perspectives on Psychological Science* **14**(4):523–42. [VFR]
Spunt, R. P., Kemmerer, D. & Adolphs, R. (2016) The neural basis of conceptualizing the same action at different levels of abstraction. *Social Cognitive and Affective Neuroscience* **11**(7):1141–51. https://doi.org/10.1093/scan/nsv084. [ZZ]
Spurrett, D. (2019) The descent of preferences. *British Journal for the Philosophy of Science*, axz020. [DS]
Stanovich, K. E. & West, R. F. (2000) Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* **23**(5):645–65. [aMG]
Stark, C., Stark, S. & Gordon, B. (2005) New semantic learning and generalization in a patient with amnesia. *Neuropsychology* **19**(2):139. [aMG]
Stawarczyk, D., Bezdek, M. A.Z & Zacks, J. M. (2019) Event representations and predictive processing: The role of the midline default network core. *Topics in Cognitive Science*, tops.12450. https://doi.org/10.1111/tops.12450. [MlS]
Steriade, D. (1995) Underspecification and markedness. In: *The handbook of phonological theory*, ed. J. A. Goldsmith, pp. 114–74. Blackwell. [MtS]
Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L. … Corlett, P. R. (2018) The predictive coding account of psychosis. *Biological Psychiatry* **84**(9):634–43. https://doi.org/10.1016/j.biopsych.2018.05.015. [PL]
Sterzer, P., Voss, M., Schlagenhauf, F. & Heinz, A. (2019) Decision-making in schizophrenia: A predictive-coding perspective. *NeuroImage* **190**:133–43. https://doi.org/10.1016/j.neuroimage.2018.05.074. [PL]
Stevens, K. N. (2002) Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America* **111**:1872–91. [MtS]
Stevenson, R. A., Baum, S. H., Segers, M., Ferber, S., Barense, M. D. & Wallace, M. T. (2017) Multisensory speech perception in autism spectrum disorder: From phoneme to whole-word perception. *Autism Research* **10**(7):1280–90. [MKB]
Stickgold, R. & Walker, M. P. (2013) Sleep-dependent memory triage: Evolving generalization through selective processing. *Nature Neuroscience* **16**(2):139–45. [AV]
Strohminger, M. & Yli-Vakkuri, J. (2018) Moderate modal skepticism. In: *Knowledge, belief, and god: New insights in religious epistemology*, ed. M. Benton, J. Hawthorne & D. Rabinowitz, pp. 302–321. Oxford University Press. [DD]
Suddendorf, T. (2013) *The gap: The science of what separates us from other animals*. Basic Books. [aMG]
Suddendorf, T. & Corballis, M. C. (2007) The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences* **30**(3):299–313. Available at: http://doi.org/10.1017/s0140525×07001975. [aMG, TP]
Suddendorf, T. & Corballis, M. C. (2010) Behavioural evidence for mental time travel in nonhuman animals. *Behavioural Brain Research* **215**(2):292–298. [TP]
Sunstein, C. R. (2003) Terrorism and probability neglect. *Journal of Risk and Uncertainty* **26**(2–3):121–36. [KM]
Sutton, R. S. & Barto, A. G. (1998) *Reinforcement learning: An introduction (adaptive computation and machine learning)*. The MIT Press. [PD]
Szpunar, K. K. & Schacter, D. L. (2013) Get real: Effects of repeated simulation and emotion on the perceived plausibility of future experiences. *Journal of Experimental Psychology: General* **142**(2):323–27. [KM]
Tamir, D. I. & Mitchell, J. P. (2010) Neural correlates of anchoring-and-adjustment during mentalizing. *Proceedings of the National Academy of Sciences of the United States of America* **107**(24):10827–832. Available at: http://doi.org/10.1073/pnas.1003242107. [aMG]
Tamir, D. I. & Mitchell, J. P. (2011) The default network distinguishes construals of proximal versus distal events. *Journal of Cognitive Neuroscience* **23**(10):2945–55. [aMG]
Tamir, D. I., Bricker, A. B., Dodell-Feder, D. & Mitchell, J. P. (2015) Reading fiction and reading minds: The role of simulation in the default network. *Social Cognitive and Affective Neuroscience* **11**(2):215–24. [aMG]
Tarhan, L. Y., Watson, C. E. & Buxbaum, L. J. (2015) Shared and distinct neuroanatomic regions critical for tool-related action production and recognition: Evidence from 131 left-hemisphere stroke patients. *Journal of Cognitive Neuroscience* **27**(12):2491–511. [rMG]
Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. (2011) How to grow a mind: Statistics, structure, and abstraction. *Science* **331**(6022):1279–85. doi: 10.1126/science.1192788. [AD, aMG]
Thomas, K., Malcolm, B. & Lastra, D. (2017) Psilocybin-assisted therapy: A review of a novel treatment for psychiatric disorders. *Journal of Psychoactive Drugs* **49**:446–55. [CF]
Thompson, C. K., Bonakdarpour, B., Fix, S. C., Blumenfeld, H. K., Parrish, T. B., Gitelman, D. R. & Mesulam, M. M. (2007) Neural correlates of verb argument structure processing. *Journal of Cognitive Neuroscience* **19**(11):1753–67. Available at: http://doi.org/10.1162/jocn.2007.19.11.1753. [aMG]
Tolman, E. C. (1948) Cognitive maps in rats and men. *Psychological Review* **55**(4):189. [aMG]
Tomasino, B., Fabbro, F. & Brambilla, P. (2014) How do conceptual representations interact with processing demands: An fMRI study on action- and abstract-related words. *Brain Research* **1591**:38–52. [rMG]
Tomasino, B. & Rumiati, R. (2013) At the mercy of strategies: The role of motor representations in language understanding. *Frontiers in Psychology* **4**:27. doi: 10.3389/fpsyg.2013.00027. [rMG]
Troche, J., Crutch, S. J. & Reilly, J. (2017) Defining a conceptual topography of word concreteness: Clustering properties of emotion, sensation, and magnitude among 750 English words. *Frontiers in Psychology* **8**(1787). doi: 10.3389/fpsyg.2017.01787 . [MP]
Trope, Y. (1986) Identification and inferential processes in dispositional attribution. *Psychological Review* **93**(3):239–57. Available at: http://doi.org/10.1037//0033-295x.93.3.239. [aMG]
Trope, Y. (1989) Levels of inference in dispositional judgement. *Social Cognition* **7** (3):296–314. [MKB]
Trope, Y. & Liberman, N. (2000) Temporal construal and time-dependent changes in preference. *Journal of Personality and Social Psychology* **79**(6):876. [aMG]
Trope, Y. & Liberman, N. (2003) Temporal construal. *Psychological Review* **110**(3):403–21. [MKB]
Trope, Y. & Liberman, N. (2010) Construal-level theory of psychological distance. *Psychological Review* **117**(2):440–63. [aMG, KFi, MR-M]
Tulving, E. (1972) Episodic and semantic memory. In: *Organization of memory*, pp. xiii, 423. Academic Press. [ZZ]
Tulving, E. (1984) Precis of Tulving elements of episodic memory. *Behavioral and Brain Sciences* **7**(2):223–38. [aMG]
Tulving, E. (1985) How many memory systems are there? *American Psychologist* **40** (4):385–98. Available at: http://doi.org/10.1037/0003-066x.40.4.385. [aMG]
Tulving, E. (1986) What kind of a hypothesis is the distinction between episodic semantic memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition* **12** (2):307–11. Available at: http://doi.org/10.1037/0278-7393.12.2.307. [aMG]
Tulving, E., Hayman, C. & Macdonald, C. A. (1991) Long-lasting perceptual priming and semantic learning in amnesia: A case experiment. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **17**(4):595. [aMG]
Turner, J. C. (1991) *Social influence*. Thomson Brooks/Cole. [aMG]
Tversky, A. (1977) Features of similarity. *Psychological Review* **84**(4):327. [aMG]
Uddin, L. Q. (2015) Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience* **16**:55–61. [CF]
Ungerleider, L. G. & Haxby, J. V. (1994) "What" and "where" in the human brain. *Current Opinion in Neurobiology* **4**(2):157–65. doi: 10.1016/0959-4388(94)90066-3. [KFr]
Vallacher, R. R. & Wegner, D. M. (1987) What do people think they're doing? Action identification and human behavior. *Psychological Review* **94**(1):3. [aMG]
Van Boven, L., Kane, J., McGraw, A. P. & Dale, J. (2010) Feeling close: Emotional intensity reduces perceived psychological distance. *Journal of Personality and Social Psychology* **98**(6):872–85. [KM]
Van Boven, L. & Loewenstein, G. (2003) Social projection of transient drive states. *Personality and Social Psychology Bulletin* **29**(9):1159–68. [KM]
Van Boven, L., Loewenstein, G., Dunning, D. & Nordgren, L. F. (2013) Changing places: A dual judgment model of empathy gaps in emotional perspective taking. In: *Advances in experimental social psychology*, vol. 48, pp. 117–71. Academic Press. [KM]

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L. & Wagemans, J. (2014) Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review* **121**(4):649–75. [MKB]

van der Meer, M. A., Johnson, A., Schmitzer-Torbert, N. C. & Redish, A. D. (2010) Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task. *Neuron* **67**(1):25–32. [aMG]

Van Kerckhove, A., Geuens, M. & Vermeir, I. (2014) The floor is nearer than the sky: How looking up or down affects construal level. *Journal of Consumer Research* **41** (6):1358–71. [KM]

Vankrunkelsven, H., Verheyen, S., Storms, G. & De Deyne, S. (2018) Predicting lexical norms: A comparison between a word association model and text-based word co-occurrence models. *Journal of Cognition* **1**(1). [MM]

Velleman, J. D. (1992) The guise of the good. *Nous* **26**(1):3–26. [aMG]

Vigliocco, G., Kousta, S. T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T. & Cappa, S. F. (2013) The neural representation of abstract words: The role of emotion. *Cerebral Cortex* **24**(7):1767–77. [MM]

Villani, C., Lugli, L., Liuzza, M. T. & Borghi, A. M. (2019) Varieties of abstract concepts and their multiple dimensions. *Language and Cognition* **11**(3):403–30. [AMB]

Walla, P. (2008) Olfaction and its dynamic influence on word and face processing: Cross-modal integration. *Progress in Neurobiology* **84**(2):192–209. [MP]

Wang, J., Conder, J. A., Blitzer, D. N. & Shinkareva, S. V. (2010) Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies. *Human Brain Mapping* **31**(10):1459–68. [arMG, MM]

Wang, L., Uhrig, L., Jarraya, B. & Dehaene, S. (2015) Representation of numerical and sequential patterns in macaque and human brains. *Current Biology* **25**(15):1966–1974. [TP]

Wang, X., Wu, W., Ling, Z., Xu, Y., Fang, Y., Wang, X., Binder, J. R., Men, W., Gao, J. H. & Bi, Y. (2018) Organizational principles of abstract words in the human brain. *Cerebral Cortex* **28**(12):4305–18. https://doi.org/10.1093/cercor/bhx283. [MlS, MM]

Wang, Y. A., da Silva Frost, A., Eastwick, P. W. & Ledgerwood, A. (2020) *Summarized attribute preferences have unique antecedents and consequences*. Unpublished Manuscript, University of California. [AL]

Watson, C. E., Cardillo, E. R., Ianni, G. R. & Chatterjee, A. (2013) Action concepts in the brain: An activation likelihood estimation meta-analysis. *Journal of Cognitive Neuroscience* **25**(8):1191–205. doi: 10.1162/jocn_a_00401. [rMG]

Weingarten, E., Chen, Q. J., McAdams, M., Yi, J., Hepler, J. & Albarracin, D. (2016) From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. *Psychological Bulletin* **142**(5):472–97. Available at: http://doi.org/10.1037/bul0000030. [aMG]

West-Eberhard, M. J. (1989) Phenotypic plasticity and the origins of diversity. *Annual Review of Ecology and Systematics* **20**:249–78. [MFe]

West-Eberhard, M. J. (2003) *Developmental plasticity and evolution*. OUP. [MFe]

Whitaker, K. J., Vendetti, M. S., Wendelken, C. & Bunge, S. A. (2018) Neuroscientific insights into the development of analogical reasoning. *Developmental Science* **21**(2): e12531. [aMG]

White, R. E. & Carlson, S. M. (2016) What would Batman do? Self-distancing improves executive function in young children. *Developmental Science* **19**(3):419–26. doi: 10.1111/desc.12314. [BO]

Williams, A., Reddigari, S. & Pylkkänen, L. (2017) Early sensitivity of left perisylvian cortex to relationality in nouns and verbs. *Neuropsychologia* **100**:131–43. [aMG]

Williams, B. (1973) Imagination and the self. In: *Problems of the self*, pp. 26–45. Cambridge University Press. [DD]

Williams, D. (2018) Hierarchical Bayesian models of delusion. *Consciousness and Cognition* **61**:129–47. https://doi.org/10.1016/j.concog.2018.03.003. [PL]

Williams, J. M. & Scott, J. (1988) Autobiographical memory in depression. *Psychological Medicine* **18**(3):689–95. https://doi.org/10.1017/s0033291700008370. [ZZ]

Williamson, T. (2007) *The philosophy of philosophy*. Blackwell. [DD]

Witkin, H. A. (1979) Socialization, culture and ecology in the development of group and sex differences in cognitive style. *Human Development* **22**:358–72. [MKB]

Witkin, H. A. & Asch, S. E. (1948) Studies in space orientation; perception of the upright in the absence of a visual field. *Journal of Experimental Psychology* **38**(5):603–14. [MKB]

Witkin, H. A., Dyk, R. B., Fattuson, H. F., Goodenough, D. R. & Karp, S. A. (1962) *Psychological differentiation: Studies of development*. Wiley. [MKB]

Witkin, H. A. & Goodenough, D. R. (1977) Field dependence and interpersonal behavior. *Psychological Bulletin* **84**(4):661–89. [MKB]

Wood, J. N. & Grafman, J. (2003) Human prefrontal cortex: processing and representational perspectives. *Nature Reviews Neuroscience* **4**(2):139–47. doi: 10.1038/nrn1033. [aMG, JG]

Xu, F. (2020) Towards a rational constructivist theory of cognitive development. *Psychological Review* **126**:841–64. http://dx.doi.org/10.1037/rev0000153. [MFe]

Xu, F. & Kushnir, T. (2012) *Rational constructivism in cognitive development, vol. 43*. Academic Press. [MFe]

Yablo, S. (1993) Is conceivability a guide to possibility? *Philosophy and Phenomenological Research* **53**:1–42. [DD]

Yee, E. (2019) Abstraction and concepts: When, how, where, what and why? *Language, Cognition and Neuroscience* **34**(10):1257–65. [CPD]

Yip-Bannicq, M. (2018) Construal enhanced conflict management: The role of abstraction in the regulation of conflict in romantic relationships [doctoral dissertation, New York University]. [MR-M]

Ylinen, A., Bragin, A., Nadasdy, Z., Jando, G., Szabo, I., Sik, A. & Buzsaki, G. (1995) Sharp wave associated high-frequency oscillation (200-hz) in the intact hippocampus: Network and intracellular mechanisms. *Journal of Neuroscience* **15**(1):30–46. [aMG]

Yudkin, D. A., Pick, R., Hur, E. Y., Liberman, N. & Trope, Y. (2019) Psychological distance promotes exploration in search of a global maximum. *Personality and Social Psychology Bulletin* **45**(6):893–906. [aMG]

Zacks, J. M., Kurby, C. A., Landazabal, C. S., Krueger, F. & Grafman, J. (2016) Effects of penetrating traumatic brain injury on event segmentation and memory. *Cortex* **74**:233–46. doi: 10.1016/j.cortex.2015.11.002. [JG]

Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A. & Hasson, U. (2017) How we transmit memories to other brains: Constructing shared neural representations via communication. *Cerebral Cortex* **27**(10):4988–5000. https://doi.org/10.1093/cercor/bhx202. [MlS]

Zarate, M. A. & Smith, E. R. (1990) Person categorization and stereotyping. *Social Cognition* **8**(2):161–85. [aMG]

Zhou, G. Y., Lane, G., Cooper, S. L., Kahnt, T. & Zelano, C. (2019) Characterizing functional pathways of the human olfactory system. *Elife* **8**. doi: ARTN e47177. 10.7554/eLife.47177. [MP]

Zimmerer, V. C., Varley, R. A., Deamer, F. & Hinzen, W. (2019) Factive and counterfactive interpretation of embedded clauses in aphasia and its relationship with lexical, syntactic and general cognitive capacities. *Journal of Neurolinguistics* **49**:29–44. Available at: http://doi.org/10.1016/j.jneuroling.2018.08.002. [aMG]

Zwaan, R. A. (1999) Situation models. The mental leap into imagined worlds. *Current Directions in Psychological Science* **8**:15–18. [DD]

Zwaan, R. A. (2016) Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin and Review* **23**:1028–1034. [DD]