KI - Künstliche Intelligenz
https://doi.org/10.1007/s13218-021-00705-x

DISCUSSION

Towards Strong AI

Martin V. Butz[1,2]

Received: 8 June 2020 / Accepted: 13 January 2021

Abstract
Strong AI—artificial intelligence that is in all respects at least as intelligent as humans—is still out of reach. Current AI lacks common sense, that is, it is not able to infer, understand, or explain the hidden processes, forces, and causes behind data. Main stream machine learning research on deep artificial neural networks (ANNs) may even be characterized as being behavioristic. In contrast, various sources of evidence from cognitive science suggest that human brains engage in the active development of compositional generative predictive models (CGPMs) from their self-generated sensorimotor experiences. Guided by evolutionarily-shaped inductive learning and information processing biases, they exhibit the tendency to organize the gathered experiences into event-predictive encodings. Meanwhile, they infer and optimize behavior and attention by means of both epistemic- and homeostasis-oriented drives. I argue that AI research should set a stronger focus on learning CGPMs of the hidden causes that lead to the registered observations. Endowed with suitable information-processing biases, AI may develop that will be able to explain the reality it is confronted with, reason about it, and find adaptive solutions, making it Strong AI. Seeing that such Strong AI can be equipped with a mental capacity and computational resources that exceed those of humans, the resulting system may have the potential to guide our knowledge, technology, and policies into sustainable directions. Clearly, though, Strong AI may also be used to manipulate us even more. Thus, it will be on us to put good, far-reaching and long-term, homeostasis-oriented purpose into these machines.

Keywords Generative predictive models · Compositionality · machine learning · Artificial neural networks · Behaviorism · Inductive learning biases · Event-predictive cognition · Causality · Homeostasis · Curiosity

1 Prolog

Another AI wave is rushing through. An event we have seen before in so many disciplines. Starting conditions are marked by surprising and partially ground-breaking successes, which are pushed by skilled protagonists. The wave is fueled by investments and hope for further revenues. Protagonists and influential companies, having built up their infrastructure, team size, and social networks, focus on both optimizing the available techniques and selling the currently best system approaches. As a result, a large part of the available intellectual power narrows down on one subject. Meanwhile, this narrowing hinders (often unintentionally) deeper innovative progress. Peer reviewing, for example, inevitably generates this side-effect.

We have seen and experienced the ceasing power of such wave-like events. The endings are typically marked by the accumulating evidence that the gained insights—the abilities of the system, the method, or the scientific approach—are not as deep and profound as originally thought. That is, the successful approach has its limits. In the AI community, the subsequent time period has been termed 'AI Winter', namely the event that is characterized by low investments, general skepticism, and a focus on other potent computational approaches. Are we heading in this direction again, seeing that the limits of the currently favored end-to-end deep learning approaches become acknowledged? Or is there potential for a sustainable, AI-supported future?

✉ Martin V. Butz
martin.butz@uni-tuebingen.de

[1] Department of Computer Science, Neuro-Cognitive Modeling Group, Faculty of Science, University of Tbingen, Sand 14, 72076 Tubingen, Germany

[2] Department of Psychology, Neuro-Cognitive Modeling Group, Faculty of Science, University of Tbingen, Sand 14, 72076 Tubingen, Germany

Published online: 26 February 2021

2 Past Reflections

With this discussion article I do not want to downscale the great recent achievements of deep learning. Nonetheless, it needs to be acknowledged that the exponential growth in computational capacity—combined with partially even faster exponential growth in data storage volume and network traffic—has enabled much of the recent success. Essentially, exponential growth has enabled us to generate more productive research on deep learning and related ANNs [93] because much more experimentation and evaluation is now possible with significantly larger networks.

The initial ground-breaker and impulse of the current ML wave was generated by Alex Krizhevsky together with Ilya Sutskever and senior AI and particularly ANN genius Geoffrey Hinton. The network, which is now simply referred to as AlexNet, busted the ImageNet competition in 2012, yielding a top-5 test error rate of 15.3%, compared to 26.3% achieved by the second-best entry. This second-best entry was still a 'traditional' approach, which used a weighted sum of scores from various types of pre-defined, feature-based classifiers. Over the next few years, the error dropped further, now reaching human-competitive or even superior top-5 test errors around 2% [86].

Several big bangs followed. Partially human-competitive performance was achieved in Atari games [73] with deep networks that develop game-critical feature encodings and consequent state-action mappings solely from reward feedback (i.e. the game score). Deep machine translation networks started to be applied by Google and others, partially outperforming traditional approaches and generating reasonable translations—even between language pairs that they had not been trained on at all [1, 116]. Finally, AlphaZero [101] has learned to play Go from scratch simply by playing against itself. It is provided with the model of the game and learns to identify game-critical, substructural patterns, which it uses to evaluate likely future game states. AlphaZero may now be considered nearly unbeatable by a human player. Even StarCraft—a real-time multi agent strategy game that hosts championships, whose games are partially broadcasted live on national TV in, for example, South Korea—was mastered by AlphaStar [113].

These results are without doubt highly impressive and should be considered great achievements in designing and training deep neural network architectures end-to-end. Success is generated by suitably designed network architectures, but without any pretraining or modular system recombination, and without explicit feature design or elaborate data preprocessing. During end-to-end training, a predefined loss signal is propagated inversely through the feed-forward processing network architecture. Direct supervised loss or reward difference signals propagate gradients back onto action outputs and further back towards the provided data input, modifying the network's weight parameters along the way. In the tasks in which planning is inevitably required, the ML algorithms are endowed with a model of the game and the ability to both anticipate future game states and to explore those states in a probabilistic, goal-oriented manner by means of rapidly exploring random tree search [36].

3 Behaviorstic Machine Learning (BML)

It is possible to draw an analogy between current AI developments and historical (but partially still ongoing) developments in psychology: the 'hype' of behaviorism. The behavioristic movement mainly succeeded because pure stimulus-response behavior was scrutinized and psychology was established as its own scientific discipline [43]. As a result, behaviorism [114] was born and it dominated psychological research in the 20th century [102].

This development may be considered somewhat surprising, seeing that many great psychologists of the time, even including William James [53], had assessed that inner states in our minds must be responsible for our goal-directed actions. Other cognitivists and linguists generated empirical evidence and argued accordingly. For example, empirical observations of adaptive behavior in rats indicated the latent learning of cognitive maps [110]. Later, language learning was suggested to proceed much faster than explainable with behavioristic theories [21]. Nonetheless, probably due to the fact that measurable results were generated easier with behavioristic paradigms—such as the infamous Skinner Box—than with cognitivist theories, behaviorism maintained its dominance over most of the twienth century.

Now entering the third decade of the twenty-first century, it seems that deep learning research is partially falling into the same trap by focusing their efforts on a paradigm, which may be called *behavioristic machine learning* (BML). When comparing these algorithms to approaches and theories in computational cognitive science and cognitive psychology (cf., e.g. [13, 14, 18, 32, 49, 54, 55]), it soon becomes apparent that current deep learning adheres to reactive, behavioristic approaches (cf., e.g. [6, 18, 62, 68]). Inputs are mapped onto target outputs, such as classifications, words of a translated sentence, or actions and reward values, optimizing the involved model parameters (i.e. weights in an ANN) to maximize target prediction accuracy. As a result, the systems act in an either fully reactive or purely reward-oriented manner, that is, they are behavioristic.

BML detects and exploits data regularities. It identifies the main tendencies and practices in the status quo, which is contained in the available data. Even when designed to solve

well-defined games, such as Go, where the ML system does look ahead, it only plans within the known data space (i.e. the game states and rules) and focuses on one static reward function [60]. Thus, even in situations when forward planning is applied, the ML algorithm only optimizes the best possible strategy within the status quo. As recommender systems, BML fosters trends and pushes towards main stream (including extremist) opinions. It focuses on identifying main data regularities, which may reflect social media trends, legislative decision making tendencies, or even correspondences in linguistic expressions. Thus, BML is data reflective rather than prospective.

The reflectively identified data regularities are very powerful, nonetheless. As detailed above, BML has shown to tremendously improve, for example, image classification accuracy, behavioral decision making in well-defined domains, and language translation systems, generating significant profit. Additionally, BML is effectively stimulating the market, particularly also via personalized advertisements, yielding even higher profit. The Zeitgeist seems to suggest: let us mine and exploit the data as best as we can, reap the profits, and see where this leads us. It is my strong hope that we can do better than that.

4 Strong AI

Related criticism about current deep learning has been raised numerous times before (cf., e.g. [6, 23, 60, 68]), albeit not directly in relation to behaviorism. Gary Marcus [68] characterized deep learning as overly data hungry with hardly any potential for transfer learning or the formation of compositional hierarchical structures. It seems unable to complete or infer hidden information, which are elsewhere referred to as 'dark' causes, that is, the causes that are not directly detectable by static visual image analysis [119]. Moreover, Marcus emphasizes that deep learning is not sufficiently transparent; it is unable to explain its decisions—in fact, it does not tend to develop explanatory decisions and is inherently not designed to discern causation from mere correlation. Furthermore, despite the best efforts over the last years, deep learning is still easily fooled [74], that is, it remains very hard to make any guarantees about how the system will behave given data that departs from the training set statistics. Finally, because deep learning does not learn causality—or generative models of hidden causes—it remains reactive, bound by the data it was given to explore [68].

In contrast, brains act proactively and are partially driven by endogenous curiosity, that is, an internal, epistemic, consistency- and knowledge-gain-oriented drive [7, 78, 91]. They develop and actively optimize predictive models, which attempt to infer the hidden causes that generate the accumulating sensorimotor experiences [33, 49, 82]. On an intuitive level, it appears that our brains attempt to predictively encode and conceptualize what is going on around us. We learn from our actively gathered sensorimotor experiences and form conceptual, loosely hierarchically structured, compositional generative predictive models, which I will refer to as *CGPMs* in the remainder of this work. Further details on CGPMs can be found in Sect. 4.2, where I scrutinize their fundamental functional and computational properties in the light of the available literature. Importantly, CGPMs allow us to reflect on, reason about, anticipate, or simply imagine scenes, situations, and developments within in a highly flexible, compositional, that is, semantically meaningful manner.[1] As a result, CGPMs enable us to actively infer highly flexible and adaptive goal-directed behavior under varying circumstances.

Seeing that we are not behavioristic automata, but humans, who reason with the help of CGPMs, I would like to suggest that AI-oriented research resources should be distributed more heterogeneously, instead of focusing them on BMLs. Ideally, AI-oriented research programs should encourage the development of techniques that promise to foster AI that learns to understand structures and interactions in our world in a conceptual, compositional manner. Such AI could issue, suggest, or recommend flexible and adaptive goal-directed actions, which, ideally, should be targeted towards a sustainable future. Due to the involved CGPMs, this AI should even be able to explain its reasoning behind its proposed recommendations. For the sake of brevity, I will refer to this kind of AI as *Strong AI* in the remainder of this article.

Strong AI has been used as a term in various disciplines and with various foci. In philosophy, John Searl has contrasted Strong AI from Weak AI, where the latter is closely related to BML [98]. The Chinese Room argument attempts to illustrate the main point: even if a machine will pass something like the Turing Test [111], it may be far from actually exhibiting a human like mind including human consciousness [99]. Particularly the qualitative experience of such a machine's 'life' will remain that of a symbol-manipulating machine. Albeit I am not addressing consciousness or qualia in this article, I put forward that the cognitive abilities of a Strong AI need to go beyond symbol manipulations.

More recently, Strong AI has been partially used as a synonym for *high-level machine intelligence*, *human-level AI*, or *general AI* [8, 42]. Partially this goes as far as the creation of a machine that is able to perform all imaginable

[1] Please note that I use the term *compositional* in a sensorimotor-grounded sense much along the lines of perceptual symbol systems [5]. As a consequence, the compositional principles that I am referring to go beyond syntactic, rule-based, or formal set-based operative compositionality [108] because they inherently integrate conceptual world knowledge.

human jobs, including all physical and all mental ones. Seeing that I am less concerned with robotics, or particular benchmark tests, here, the closest relation to Strong AI, as I use the term, may be drawn to *Cognitive AI* [119]—AI that can develop common sense reasoning abilities [23, 62, 64, 70, 72].

I propose that, in order to develop such Strong AI, we need systems that are able to learn CGPMs of their encountered environment. With the help of CGPMs and suitable inference processes, Strong AI will be able to reason about its environment. It will exhibit common sense, because it will be able to identify, reason about, and explicate causal relations. Moreover, it will be able to act upon—or propose actions within—its encountered environment in a goal- and value-oriented manner, pursuing both knowledge gain and homeostasis. Clearly, numerous questions on how to create such Strong AI remain wide open:

- Learning conceptual structures: How may the conceptualizations in CGPMs be learned?
- Discerning causality: How can the critical hidden causal aspects of the processes and forces behind our observations be learned?
- World-knowledge-grounded compositionality: How can learned conceptualizations be combined in seemingly infinite compositionally meaningful manners?
- Compositional reasoning and decision making: How can compositional knowledge structures be used to plan ahead in a highly adaptive and flexible goal- or value-oriented manner?

Before I address these questions in the next section, one possible concern should be addressed: we humans tend to make mistakes, we sometimes develop false beliefs and superstitions, and we often do not succeed in taking all relevant factors into account when making decisions (or when optimizing behavior, more generally speaking). Some of these failures can be explained by our tendency to develop heuristics and habits, many of which have actually been shown to be relatively effective [39, 40]. Other types of failures cannot be directly related to heuristics-based reasoning. Rather, these deficits can be explained by resource limitations in our brains, as suggested by the success of resource-rational cognitive modeling approaches [65]. This also implies that more resources may enable deeper rationality, diminishing the present human deficits. Thus, the types of CGPMs that we humans are able to learn, as well as the reasoning mechanisms that we use to exploit CGPMs to make good decisions, appear to be very much worth pursuing when aiming at developing Strong AI; particularly when this Strong AI is equipped with a sufficiently large amount of computational resources.

5 Inductive Learning and Processing Biases

The development of truly intelligent, Strong AI seems to be only possible if we employ the right inductive processing and learning biases to enable the learning of CGPMs [6, 14, 15, 62]. When considering brain development and cognition, it has become obvious that evolution has equipped us with numerous such inductive biases to maximize our chances of survival on an evolutionary scale [24]. Simply put, it appears that evolution has discovered that CGPMs enable the pursuance of more social, adaptive, versatile, and anticipatory goal-directed behavior [16]. From a more cognitive perspective it may be said that CGPMs enable us to reason and ask questions in an interventional, prospective as well as in a counterfactual, memorizing, and consolidating, retrospective manner [79, 80]. Furthermore, effective compositionality allows us to do so in an analogical, innovative manner, enabling zero-shot learning, that is, to act effectively under circumstances that are only loosely related to previous situations.

In line with these cognitive science-based suggestions, the current deep learning successes essentially also show that hard-coded features are typically not as effective as a rather open-ended feature processing architecture. Generative models are extremely hard to pre-structure in a hard-coded manner. Our world is simply too complex. Instead, as Rich Sutton has put it in his thoughts on "The Bitter Lesson": "[...] we should build in only the meta-methods that can find and capture this arbitrary complexity. [That is,] We want AI agents that can discover like we can, not which contain what we have discovered." [105, p.1]. A similar argument can be put forward from a pure optimization perspective: the *No-Free-Lunch theorem* clearly implies that some biases are needed to optimize learning in environments that adhere to particular principles, like space, time, energy, or matter [12, 115].

Accordingly, I argue that we need to equip our ML systems with suitable inductive learning and processing biases to foster the active construction of CGPMs. More particularly, I will put forward that one important type of inductive learning bias may lie in the tendency to construct event-predictive encodings and abstractions thereof. Moreover, the learning systems should be open-ended. Thus, reasoning, planning, and behavioral control should incorporate an inductive processing bias that maintains a healthy balance between epistemic, that is, knowledge gain-oriented, and homeostasis-oriented behavior. As a result, experience-grounded CGPMs will be effectively learned and exploited, while exploring and manipulating the encountered environment. This environment may be our actual world, which may be explored with a robot [37, 67] or an agentive system, which could also interact with a simulated reality.

5.1 Generative Predictive Models

Generative predictive models (GPMs), as characterized in this section, are fundamentally different from BML because they do not learn conditional classifications or behavioral patterns, given data. Rather, they develop joint probabilities, generally speaking. Moreover, they should be temporally predictive, in that they are attempting to learn the processes and forces behind the causes that generated the observable data. GPMs should not be confused with Generative Adversarial Networks (GANs). GANs combine an encoder network, the predictor or classifier, with a decoder network, which generates data. Although any decoder network may be considered to be 'generative', GANs are designed to generate data patterns that challenge the encoder.

GPMs are closely related to predictive coding [83], the predictive brain [10, 22], and generative perception [44, 95, 96]. They are most generally formulated in Karl Friston's Free Energy principle [31–33]. In short, the formalism implies that brains attempt to minimize anticipated uncertainty about both future sensory impressions and inner states, where the latter should not diverge from homeostasis[2]. More specifically, it implies that brains attempt to (i) know what is going on, (ii) learn from experience, and (iii) pursue epistemic and homeostasis-oriented behavior: Retrospective, rather fast updating of generative model activities yields latent state hypotheses about the current—but also hypothetical other—states of affairs. Slower adaptive processes, which selectively integrate more experience, learn and consolidate knowledge by adapting the parameters of the developing generative model. Finally, active, prospective inference triggers motor activities that are believed to minimize anticipated future surprises, yielding epistemic, goal-oriented behavior [34]. These computational cognitive modeling principles also imply that they can be implemented in deep ANN architectures [17, 51, 76, 77].

5.2 Compositional Generative Predictive Models

While GPMs are certainly useful, they are even more powerful when they can be learned fast, use little energy-related resources, and are maximally suited to generate adaptive behavior. *Compositional* GPMs (i.e., CGPMs), as I refer to them here, encode conceptual, hierarchical, causal models, which enable the recombination of GPM components in semantically meaningful, world knowledge-grounded manners. Various researchers have emphasized the importance of *compositionality*, which is essentially hardly if at all developing within current deep learning approaches [6, 14, 62, 68].

One important ingredient for developing compositional structures is a solution to the binding problem [18, 94], that is, the problem to flexibly bind features—of whatever kind generally speaking—into coherent wholes. This solution must be realized by some form of neural dynamics that are able to selectively integrate multiple features into a consistent, overall structure. Given that features are encoded predictively, the activation of features inherently activates predictions of the activities of other features, besides predictions about actual sensory impressions. As a result, coherence in the active structure may be measured by the resulting mutual prediction error. Gregor Schöner's dynamic neural field theory mimics such a mechanism: neural competitive dynamics fall into integrative, distributed neural attractors, where the activities in the involved modularized feature spaces condition each other in a predictive manner [88, 97]. In my own group, Fabian Schrodt has shown that an effective combination of autoencoder-based GPMs and redundant, distributed, population-based feature encodings enables Gestalt inference [89, 95]. In this case, the internal perspective is adjusted while biological motion features are flexibly bound into Gestalt percepts, given that known patterns can be detected.

The compositionality-oriented challenge to generate dynamic trajectories, to, for example, learn to both recognize and draw letters and other symbols has been considered [61]. Seeing that humans are very fast in learning new symbols—essentially in a one-shot manner—compositional recombinations of dynamic sub-trajectories appear to be at hand [29, 59, 60]. We have recently shown that a suitably-structured recurrent ANN architecture can yield similar compositional structures, that is, a sensorimotor-grounded CGPM [28]. All of these approaches are essentially able to flexibly bind and recombine sub-trajectories, thus enabling one-shot learning and innovative, compositional recombinations of, in this case, letter sub-trajectories.

For cognition in general, though, more complex components need to be compositionally bindable. These components may be related to causality, physics, functionality, intentionality, and utility, which have been identified as five key domains for a Cognitive AI elsewhere [119]. Albeit an approximate causal understanding of our world lies at the core of cognition, causal learning [11] is particularly challenging because it is very difficult to distinguish mere correlations from actual causal interactions. Intuitive physics (cf., e.g. [62]) and a functionality-oriented perception (in the sense of affordances [38]) characterize entities and potential interactions with and between them. When perceiving actual agents, intuitive

[2] Particularly the latter case prevents the system from preferring to live in a dark room, where it will inevitably starve at some point. But also the former notion generates an epistemic drive toward increasing certainty of the state estimates of the (relevant) surrounding environment.

psychology comes into play as well. Intentionality needs to be inferred to make sense of the behavior of others—a concept that is closely related to inverse planning and inverse reinforcement learning [3, 47, 87]. Finally, the concept of utility needs to be integrated, including motivations of other agents, efforts involved, as well as other negative rewards, such as when (potentially) getting hurt. All five core domains are termed 'dark' [119] in the sense that they are not directly observable. Humans clearly have a rather good grasp on them and are indeed able to combine them in a compositional manner: We are able to flexibly bind interacting entities and infer the involved hidden causes and forces that determine the entities' behaviors. We are even able to infer the knowledge and utility-originating intentions of the involved agents from a rather young age onwards [2, 41].

Learning CGPMs, which essentially need to be able to both develop such conceptual and compositionally recombinable components and bind the components in goal-directed or value-oriented manners, remains a hard nut to crack. To succeed, inductive learning biases appear necessary [6, 14] to guide CGPM development. Moreover, active hypothesis testing, that is, epistemic behavior seems necessary to be able to identify actual causality. In the remainder of this section, I suggest particular inductive learning and processing biases, which may be very useful for developing CGPMs.

5.3 Critical Inductive Learning and Processing Biases

Inductive biases help to bootstrap learning, guiding it in the right direction. Evolution clearly has encoded many such biases into our genes. Our bodies grow in a very systematic, predetermined manner. Concurrently, our brain grows, forming and continuously consolidating computational modules in a highly systematic but plastic manner. This suggests that developmental ML systems should also be endowed with general, inductive learning biases, that is, meta-methods, which suitably guide the learning process under useful structural assumptions [6, 15].

Accumulating research in cognitive science and related disciplines suggests the presence of at least two fundamental inductive biases. First, *event-predictive inductive learning biases* foster the development of loosely hierarchically structured, event-predictive models [4, 14–16, 30, 81, 84, 100]. Second, a motivational system maintains a healthy but complex balance between *epistemic- and homeostasis-oriented inductive processing biases* [19, 25, 69, 78, 90, 92].

5.3.1 Event-Predictive Inductive Learning Bias

Various strands of research in cognitive science emphasize that we perceive and act upon our world in the form of events [4, 14, 16, 58]. Given our CGPM perspective, event-predictive cognition emphasizes that events are flexibly constructed in a compositional manner. Events characterize a static or dynamic situation, in which interactions unfold systematically and predictably. They can be typically marked by a beginning, and associated constraining conditions, which enable the commencement of an event. They are furthermore characterized by typical final conditions, which often coincide with a goal and which mark the end of an event. A simple example is to grasp a glass and to drink out of it. This overall event can be partitioned into a reach, a grasp, a suitable transport to the mouth, actually drinking, and typically transporting the glass back and releasing it.

While event transitions may be more fluid in many other circumstances, it appears that our brain has a strong tendency, that is, an inductive learning bias, to segment and compress the continuous stream of sensorimotor experiences into event-predictive encodings [16]. Such event encodings have been characterized as common codes of actions and their effects (Theory of Event Coding, [50]). Moreover, they have been characterized as higher-level codes, which we utilize to segment and interpret our perceptions, but also to guide our actions and thoughts [81, 84, 103, 117, 118]. During communication, speakers encode events in utterances. Peter Gärdenfors [35] went as far as explicitly stating that "sentences express events" (p. 107), including *stative events* and *dynamic events*. Events may thus be described by a sentence, but they certainly exist independent of the particular sentence used to describe them in a conceptual, compositional, world-knowledge-grounded format [26, 27, 52, 56, 57, 71, 112].

I have previously proposed that events consist of spatial-relational encodings of entities and the forces that are played out by them and between them over the duration of the considered event [14, 15]. The development of such predictive encodings can be bootstrapped from our own sensorimotor experiences, as suggested by the mirror neuron system [85]. During development, motor commands need to be abstracted into conceptual encodings, which predict the effects of forces onto our environment. While observing the environment then, these encodings enable the inference of both the forces and the natural or agentive causes, which induced the forces in the first place. In the case of agentive causes, additionally, preferences, intentions, and even the knowledge state of the observed agent can be inferred [2, 41]. Moreover, the concept of forces can generalize away from actual physical ones enabling analogical thinking [5, 63]. 'Social pressure' or 'political

influence' are good examples. Meanwhile, the involved entities and forces may be characterized and individualized further.

A particular event in our brain is thus imagined by the active subset of all available predictive encodings in our CGPM. This subset characterizes the event's properties—possibly with rather many details about a concrete scene or scenario, but in other circumstances possibly also in a rather abstract form. Critically, though, the subset needs to form a predictive attractor, where the involved—partially mutually—predictive encodings form a local free energy minimum (that is, simplistically speaking, a local mutual prediction error minimum). As a result, the involved CGPM components are temporally bound together into a dynamic, relational code. For example, when grasping a glass in order to drink from it, hand, mouth, glass, their (approximate) spatial relation, grasp motions, sensory feedback anticipations, fluid expectations, etc. are integrated into such a predictive attractor. Event-predictive encodings may thus be viewed as attractors in an interactive network of dependencies.

In order to develop such event-predictive encodings by continuously analyzing the sensorimotor stream of environmental interactions, I propose that the key inductive learning bias is the expectation of temporally stable attractors, which encode events. Temporal instabilities mark transitions between attractors and are harder to predict (cf. the early model of Jeff Zacks [117] and related propositions elsewhere [4, 14, 16, 58, 100]). Measures of surprise have been proposed and implemented to quickly identify transitions between events, segmenting the stream of information and consolidating event codes [20, 45, 117]. Developing latent codes characterize individual events, predictively encoding typical activities and activity dynamics [45, 100] in a semantically-meaningful, compositionally recombinable manner. Vector spaces have been recently proposed to be well-suited for such encodings [30]. However, we also find potential in suitably modularized neural networks that are endowed with retrospectively inferable latent states [17, 51, 104]. Over time, event-predictive encodings develop, which predict the characteristic temporally stable dynamics that typically unfold during the event as well as conditions for the event to commence, to continue to apply, and to end.

Applied at different levels of abstraction and with different sensitivity rates, loosely hierarchically structured CGPMs can develop [46]. Note the close relation to the options framework in hierarchical reinforcement learning—a key aspect of RL that still is somewhat under-appreciated [9, 106, 107]. I thus propose that event-oriented segmentations and retrospective optimizations and consolidations very likely offer the inductive learning biases needed to develop loosely hierarchically-structured CGPMs.

5.3.2 Epistemic- and Homeostasis-Oriented Processing

The free energy-based active inference mechanism detailed by Karl Friston et al. [34] includes two optimization summands, which essentially constitute the loss function for inferring goal-directed behavior. One of them focuses on minimizing expected entropy, that is, uncertainty in the anticipated future. The other one aims at pursuing internal homeostasis. As a result, behavior is a blend between epistemic- and homeostasis-oriented processes, which activate actions and action routines in an inverse manner. A good balance between the two processes and the maintenance of this balance over time is part of this overall inductive processing bias towards knowledge gain and homeostasis [34, 109]. Interactions between the two measures due to expected uncertainties seem to be important and clearly observable in human behavior, including epistemic top-down attention [4, 48, 66].

The hierarchical structures that develop from event-predictive inductive learning biases enable us to progressively consider and optimize behavior further into the future. The epistemic bias will lead to hypothesis testing, that is, the focused generation of experiences. Playing in children is essentially acted out curiosity in imaginary scenes and events. The consequent active development of CGPMs enables the direct disambiguation of causal influences from mere correlative sensory signals. And this curiosity-driven process seems to be played-out not only during own experimentation, but also while watching and interacting with others. Meanwhile, the homeostasis-driven influences direct our attention and behavior to those aspects that are deemed relevant, because they are experienced as rewarding. For example, social interaction rewards play an important role in developing our social competence. As a result, driven by epistemic- and homeostasis-driven processing biases, CGPMs will emerge that approximate causality and focus on the aspects that are deemed relevant for one's own self.

6 Final Discussion

In this paper, I have argued that the current AI hype may be termed a Behavioristic Machine Learning (BML) wave. It is the involved blind, reactive development that I consider as unsustainable, even if short-term rewards are generated. I have suggested that research efforts should be increased to develop Strong AI, that is, artificial systems that are able to learn about the processes, forces, and causes underlying the perceived data, becoming able to understand and explain them. As a precursor, the field should target the development of world-knowledge-grounded compositional, generative predictive models (CGPMs). The development of this type of compositionality will be possible if machine learning

algorithms are enriched with suitable learning and processing biases. Event-predictive inductive learning as well as epistemic- and homeostasis-oriented inductive processing may constitute two of these biases.

CGPMs will be immensely important for the development of explainable AI, because explanations are about how things work in the world, that is, explanations are about causality. Moreover, CGPMs will be extremely useful to reason and plan in a more versatile and adaptive manner. Generally, GPMs enable retrospective consolidation, counterfactual and hypothetical reasoning and imagination, and prospective, interventional thinking [80]. On top of that, a compositional GPM structure will enable the application of the gathered knowledge under different circumstances, promising to solve hard challenges, such as zero-shot learning tasks as well as related analogical reasoning and problem solving tasks.

In conclusion, I have put forward that the development of CGPMs by means of suitable inductive learning and processing biases may pave the way for the development of Strong AI. Progress towards Strong AI is currently hindered by a lack of data (about processes and systems), by limitations in the available simulation platforms, by hardware constraints in robotics, and by the current BML focus. These obstacles will be circumvented earlier if we manage to broaden our ML and AI research efforts. Eventually, we will witness artificial systems that can reason about their actions or action propositions and explain them. Equipped with sufficient processing resources, this Strong AI will have extremely high potential. On the negative side, it may be used in a profit- or power-oriented manner to control and manipulate us far beyond current applications [75]—a development, which clearly must be avoided. On the positive side, it may support and guide us in creating an environment that is enjoyable, that satisfies our human as well as other species' needs, and that can be sustained for centuries to come. To make this happen, it will be on us to put good, far-reaching and long-term, homeostasis-oriented purpose into these Strong AI machines [87].

Acknowledgements Funding from a Feodor Lynen Research Fellowship of the Humboldt Foundation is acknowledged. Moreover, funding was received from the German Research Foundation (DFG) (Research Training Group 1808: Ambiguity—Production and Perception, project number 198647426 and project number BU 1335/11-1 in the framework of the SPP program "The Active Self", SPP 2134). Additional support comes from the DFG Cluster of Excellence "Machine Learning—New Perspectives for Science", EXC 2064/1, project number 390727645.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

1. Bahdanau D, Cho K, Bengio Y (2015) Neural machine translation by jointly learning to align and translate. In: 3rd international conference on learning representations, ICLR 2015 (2015). ArXiv:1409.0473
2. Baker CL, Jara-Ettinger J, Saxe R, Tenenbaum JB (2017) Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. Nat Hum Behav 1(4):0064. https://doi.org/10.1038/s41562-017-0064
3. Baker CL, Saxe R, Tenenbaum JB (2009) Action understanding as inverse planning. Cognition 113(3):329–349. https://doi.org/10.1016/j.cognition.2009.07.005
4. Baldwin DA, Kosie JE (2020) How does the mind render streaming experience as events? Topics in Cognitive Science. https://doi.org/10.1111/tops.12502
5. Barsalou LW (1999) Perceptual symbol systems. Behav Brain Sci 22:577–600
6. Battaglia PW, Hamrick JB, Bapst V, Sanchez-Gonzalez A, Zambaldi V, Malinowski M, Tacchetti A, Raposo D, Santoro A, Faulkner R, Gulcehre C, Song F, Ballard A, Gilmer J, Dahl G, Vaswani A, Allen K, Nash C, Langston V, Dyer C, Heess N, Wierstra D, Kohli P, Botvinick M, Vinyals O, Li Y, Pascanu R (2018) Relational inductive biases, deep learning, and graph networks. ArXiv:1806.01261
7. Berlyne DE (1960) Conflict, arousal, and curiosity. McGraw-Hill, New York
8. Besold T, Hernndez-Orallo J, Schmid U (2015) Can machine intelligence be measured in the same way as human intelligence?. Künstliche Intelligenz 29:291–297. https://doi.org/10.1007/s13218-015-0361-4
9. Botvinick M, Niv Y, Barto AC (2009) Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. Cognition 113(3):262–280. https://doi.org/10.1016/j.cognition.2008.08.011
10. Bubic A, Von Cramon DY, Schubotz RI (2010) Prediction, cognition and the brain. Front Hum Neurosci. https://doi.org/10.3389/fnhum.2010.00025
11. Buehner MJ, Chenk PW (2012) Causal learning. In: Holyoak KJ, Morrison RG (eds) The Cambridge handbook of thinking and reasoning. Cambridge University Press, Cambridge, pp 143–168
12. Butz MV (2004) Anticipation for learning, cognition, and education. Horizon 12:111–116
13. Butz MV (2008) How and why the brain lays the foundations for a conscious self. Constr Found 4(1):1–42
14. Butz MV (2016) Towards a unified sub-symbolic computational theory of cognition. Front Psychol 7:925. https://doi.org/10.3389/fpsyg.2016.00925
15. Butz MV (2017) Which structures are out there? Learning predictive compositional concepts based on social sensorimotor explorations. In: Metzinger TK, Wiese W (eds) Philosophy and predictive processing. MIND Group, Frankfurt a. M. https://doi.org/10.15502/9783958573093

16. Butz MV, Achimova A, Bilkey D, Knott A (2020) Event-predictive cognition: A root for conceptual human thought. Topics Cognitive Sci. https://doi.org/10.1111/tops.12522
17. Butz MV, Bilkey D, Humaidan D, Knott A, Otte S (2019) Learning, planning, and control in a monolithic neural event inference architecture. Neural Netw 117:135–144. https://doi.org/10.1016/j.neunet.2019.05.001
18. Butz MV, Kutter EF (2017) How the mind comes into being: Introducing cognitive science from a functional and computational perspective. Oxford University Press, Oxford
19. Butz MV, Shirinov E, Reif KL (2010) Self-organizing sensorimotor maps plus internal motivations yield animal-like behavior. Adapt Behav 18(3–4):315–337
20. Butz MV, Swarup S, Goldberg DE (2004) Effective online detection of task-independent landmarks. IlliGAL report 2004002, Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana-Champaign
21. Chomsky N (1959) Review of B. F. Skinner, Verbal Behavior. Language 35:26–58
22. Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 36:181–253
23. Davis E, Marcus G (2015) Commonsense reasoning and commonsense knowledge in artificial intelligence. Commun ACM 58(9):92–103. https://doi.org/10.1145/2701413
24. Dawkins R (1976) The Selfish Gene. Oxford University Press, Oxford
25. Dayan P, Balleine BW (2002) Reward, motivation and reinforcement learning. Neuron 36:285–298
26. Elman JL, McRae K (2019) A model of event knowledge. Psychol Rev 126:252–291. https://doi.org/10.1037/rev0000133
27. Evans V (2015) Whats in a concept? Analog versus parametric concepts in LCCM theory. In: Margolis E, Laurence S (eds) The conceptual mind: new directions in the study of concepts, chap. 10. MIT Press, Cambridge, pp 251–290
28. Fabi S, Otte S, Wiese JG, Butz MV (2020) Investigating efficient learning and compositionality in generative lstm networks. In: Farkaš I, Masulli P, Wermter S (eds) Artificial neural networks and machine learning—ICANN 2020. Springer International Publishing, Cham, pp 143–154
29. Feinman R, Lake BM (2020) Generating new concepts with hybrid neuro-symbolic models. In: Proceedings of the 42nd annual meeting of the cognitive science society, pp 2315–2321
30. Franklin NT, Norman KA, Ranganath C, Zacks JM, Gershman SJ (2020) Structured event memory: a neuro-symbolic model of event cognition. Psychol Rev 127(3):327–361. https://doi.org/10.1037/rev0000177
31. Friston K (2003) Learning and inference in the brain. Neural Netw 16(9):1325–1352. https://doi.org/10.1016/j.neunet.2003.06.005
32. Friston K (2009) The free-energy principle: a rough guide to the brain? Trends Cognit Sci 13(7):293–301. https://doi.org/10.1016/j.tics.2009.04.005
33. Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11:127–138. https://doi.org/10.1038/nrn2787
34. Friston K, Rigoli F, Ognibene D, Mathys C, FitzGerald T, Pezzulo G (2015) Active inference and epistemic value. Cognit Neurosci 6:187–214. https://doi.org/10.1080/17588928.2015.1020053
35. Gärdenfors P (2014) The geometry of meaning: semantics based on conceptual spaces. MIT Press, Cambridge
36. Gelly S, Silver D (2011) Monte-Carlo tree search and rapid action value estimation in computer Go. Artif Intell 175(11):1856–1875. https://doi.org/10.1016/j.artint.2011.03.007
37. Georgie YK, Schillaci G, Hafner VV (2019) An interdisciplinary overview of developmental indices and behavioral measures of the minimal self. In: International conference on development and learning and EpigeneticRobotics (ICDL-EpiRob), pp 129–136
38. Gibson JJ (1979) The ecological approach to visual perception. Lawrence Erlbaum Associates, Mahwah
39. Gigerenzer G, Gaissmaier W (2011) Heuristic decision making. Ann Rev Psychol 62(1):451–482. https://doi.org/10.1038/s41562-017-0064
40. Gigerenzer G, Todd PM (1999) Simple heuristics that make us smart. Oxford University Press, New York
41. Gopnik A, Wellman HM (2012) Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. Psychol Bull 138(6):1085–1108. https://doi.org/10.1038/s41562-017-0064
42. Grace K, Salvatier J, Dafoe A, Zhang B, Evans O (2018) Viewpoint: when will ai exceed human performance? evidence from ai experts. J Artif Intell Res. https://doi.org/10.1038/s41562-017-0064
43. Graham G (2019th) Behaviorism. In: Zalta EN (ed) The Stanford encyclopedia of philosophy, spring, 2019th edn. Stanford University, Metaphysics Research Lab, Stanford
44. Gross HM, Volker S, Torsten S (1999) A neural architecture for sensorimotor anticipation. Neural Netw 12:1101–1129
45. Gumbsch C, Butz MV, Martius G (2019) Autonomous identification and goal-directed invocation of event-predictive behavioral primitives. IEEE Trans Cognitive Dev Syst. https://doi.org/10.1038/s41562-017-0064
46. Gumbsch C, Otte S, Butz MV (2017) A computational model for the dynamical learning of event taxonomies. In: Proceedings of the 39th annual meeting of the cognitive science society, pp 452–457. Cognitive science society
47. Hadfield-Menell D, Russell SJ, Abbeel P, Dragan A (2016) Cooperative inverse reinforcement learning. In: Lee DD, Sugiyama M, Luxburg UV, Guyon I, Garnett R (eds.) Advances in neural information processing systems, vol 29, pp 3909–3917. Curran Associates, Inc
48. Hayhoe MM, Shrivastava A, Mruczek R, Pelz JB (2003) Visual memory and motor planning in a natural task. J Vis 3(1):49–63
49. Hoffmann J (1993) Vorhersage und Erkenntnis: Die Funktion von Antizipationen in der menschlichen Verhaltenssteuerung und Wahrnehmung. [Anticipation and cognition: the function of anticipations in human behavioral control and perception.]. Hogrefe, Göttingen
50. Hommel B, Müsseler J, Aschersleben G, Prinz W (2001) The theory of event coding (TEC): a framework for perception and action planning. Behav Brain Sci 24:849–878
51. Humaidan D, Otte S, Butz MV (2020) Fostering event compression using gated surprise. In: Farkaš I, Masulli P, Wermter S (eds) Artificial neural networks and machine learning—ICANN 2020. Springer International Publishing, Cham, pp 155–167
52. Jackendoff R (2002) Foundations of language. Brain, meaning, grammar, evolution. Oxford University Press, Oxford
53. James W (1890) The principles of psychology. Dover Publications, New York
54. Johnson-Laird PN (1983) Mental models: towards a cognitive science of language, inference, and consciousness. Cambridge University Press and Harvard University Press, Cambridge
55. Knauff M (2013) Space to reason. A spatial theory of human thought. MIT Press, Cambridge
56. Knott A (2012) Sensorimotor cognition and natural language syntax. MIT Press, Cambridge
57. Knott A, Takac M (2020) Roles for event representations in sensorimotor experience, memory formation, and language processing. Topics Cognitive Sci. https://doi.org/10.1038/s41562-017-0064

58. Kuperberg GR (2020) Tea with milk? A hierarchical generative framework of sequential event comprehension. Topics Cognitive Sci. https://doi.org/10.1038/s41562-017-0064
59. Lake BM (2019) Compositional generalization through meta sequence-to-sequence learning. In: Wallach H, Larochelle H, Beygelzimer A, dAlché-Buc F, Fox E, Garnett R (eds) Advances in neural information processing systems 32, pp 9791–9801. Curran Associates, Inc
60. Lake BM, Salakhutdinov R, Tenenbaum JB (2015) Human-level concept learning through probabilistic program induction. Science 350(6266):1332–1338. https://doi.org/10.1038/s41562-017-0064
61. Lake BM, Salakhutdinov R, Tenenbaum JB (2019) The omniglot challenge: a 3-year progress report. Curr Opin Behav Sci 29:97–104. https://doi.org/10.1038/s41562-017-0064
62. Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ (2017) Building machines that learn and think like people. Behav Brain Sci. https://doi.org/10.1038/s41562-017-0064
63. Lakoff G, Johnson M (1980) Metaphors we live by. The Universty of Chicago Press, Chicago
64. Levesque HJ (2017) Common sense, the Turing test, and the quest for real AI. MIT Press, Cambridge
65. Lieder F, Griffiths TL (2020) Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. Behav Brain Sci 43:e1. https://doi.org/10.1038/s41562-017-0064
66. Lohmann J, Belardinelli A, Butz MV (2019) Hands ahead in mind and motion: active inference in peripersonal hand space. Vision 3:2. https://doi.org/10.1016/j.cognition.2009.07.005
67. Lungarella M, Metta G, Pfeifer R, Sandini G (2003) Developmental robotics: a survey. Connect Sci 15(4):151–190. https://doi.org/10.1016/j.cognition.2009.07.005
68. Marcus G (2018) Deep learning: a critical appraisal. CoRR **abs/1801.00631**
69. Maturana H, Varela F (1980) Autopoiesis and cognition: the realization of the living. Reidel, Boston
70. McCarthy J (1959) Programs with common sense. In: Proceedings of the Teddington conference on the mechanization of thought processes. Her Majesty's Stationary Office, London, pp 75–91
71. McRae K, Brown KS, Elman JL (2019) Prediction-based learning and processing of event knowledge. Topics Cognitive Sci. https://doi.org/10.1016/j.cognition.2009.07.005
72. Minsky M (2006) The emotion machine: commonsense thinking, artificial intelligence, and the future of the human mind. Simon and Schuster
73. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, Petersen S, Beattie C, Sadik A, Antonoglou I, King H, Kumaran D, Wierstra D, Legg S, Hassabis D (2015) Human-level control through deep reinforcement learning. Nature 518(7540):529–533. https://doi.org/10.1038/nature14236
74. Nguyen A, Yosinski J, Clune J (2015) Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 427–436
75. O'Neil C (2016) Weapons of math destruction: how big data increases inequality and threatens democracy. Broadway Books
76. Otte S, Hofmaier L, Butz MV (2018) Integrative collision avoidance within rnn-driven many-joint robot arms. Artif Neural Netw Mach Learn ICANN 2018(11141):748–758
77. Otte S, Schmitt T, Friston K, Butz MV (2017) Inferring adaptive goal-directed behavior within recurrent neural networks. In: 26th international conference on artificial neural networks (ICANN17) pp 227–235
78. Oudeyer PY, Kaplan F, Hafner VV (2007) Intrinsic motivation systems for autonomous mental development. IEEE Trans Evolut Comput 11:265–286. https://doi.org/10.1016/j.cognition.2009.07.005
79. Pearl J (2000) Causality. Models, reasoning, and inference. Cambridge University Press, New York
80. Pearl J (2020) The limitations of opaque learning machines. In: Brockman J (ed) Possible minds: 25 ways of looking at AI, chap. 2. Penguin Press, New York, pp 13–19
81. Radvansky GA, Zacks JM (2014) Event cognition. Oxford University Press, Oxford
82. Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2(1):79–87. https://doi.org/10.1016/j.cognition.2009.07.005
83. Rao RPN, Ballard DH (1997) Dynamic model of visual recognition predicts neural response properties in the visual cortex. Neural Comput 9:721–763
84. Richmond LL, Zacks JM (2017) Constructing experience: event models from perception to action. Trends Cognitive Sci 21(12):962–980. https://doi.org/10.1016/j.tics.2017.08.005
85. Rizzolatti G, Sinigaglia C (2010) The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. Nat Rev Neurosci 11(4):264–274. https://doi.org/10.1016/j.cognition.2009.07.005
86. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) ImageNet large scale visual recognition challenge. Int J Comput Vis (IJCV) 115(3):211–252. https://doi.org/10.1016/j.cognition.2009.07.005
87. Russell S (2020) The purpose put into the machine. In: Brockman J (ed) Possible minds: 25 ways of looking at AI, chap. 3. Penguin Press, New York, pp 20–32
88. Sabinasz D, Richter M, Lins J, Schner G (2020) Speaker-specific adaptation to variable use of uncertainty expressions. In: Proceedings of the 42nd annual meeting of the cognitive science society, pp 620–627
89. Sadeghi M, Schrodt F, Otte S, Butz MV (2020) Binding and perspective taking as inference in a generative neural network model
90. Schillaci G, Villalpando AP, Hafner VV, Hanappe P, Colliaux D, Wintz T (2020) Intrinsic motivation and episodic memories for robot exploration of high-dimensional sensory spaces. Adapt Behav. https://doi.org/10.1016/j.cognition.2009.07.005
91. Schmidhuber J (1991) A possibility for implementing curiosity and boredom in model-building neural controllers. In: Meyer JA, Wilson SW (eds) Proceedings of the international conference on simulation of adaptive behavior, pp 222–227. MIT Press/Bradford Books
92. Schmidhuber J (1991) A possibility for implementing curiosity and boredom in model-building neural controllers. In: Proceedings of the first international conference on simulation of adaptive behavior: from animals to animats, pp 222–227
93. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117. https://doi.org/10.1111/tops.12502
94. Schmidt T (2009) Perception: the binding problem and the coherence of perception. In: Banks WP (ed) Encyclopedia of consciousness. Academic Press, Oxford, pp 147–158. https://doi.org/10.1111/tops.12502
95. Schrodt F (2018) Neurocomputational principles of action understanding: perceptual inference, predictive coding,and embodied simulation. Ph.D. thesis, Faculty of Science, University of Tbingen. https://doi.org/10.15496/publikation-24327

96. Schrodt F, Butz MV (2016) Just imagine! learning to emulate and infer actions with a stochastic generative architecture. Front Robot AI. https://doi.org/10.1111/tops.12502
97. Schner G (2019) The dynamics of neural populations capture the laws of the mind. Topics Cognitive Sci. https://doi.org/10.1111/tops.12502
98. Searle JR (1980) Minds, brains, and programs. Behav Brain Sci 3(03):417–424
99. Searle JR (1999) Chinese room argument. In: Wilson RA, Keil FC (eds) The MIT encyclopedia of the cognitive sciences. MIT Press, Cambridge, pp 115–116
100. Shin YS, DuBrow S (2020) Structuring memory through inference-based event segmentation. Topics Cognitive Sci. https://doi.org/10.1111/tops.12502
101. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S, Grewe D, Nham J, Kalchbrenner N, Sutskever I, Lillicrap T, Leach M, Kavukcuoglu K, Graepel T, Hassabis D (2016) Mastering the game of Go with deep neural networks and tree search. Nature 529(7587):484–489. https://doi.org/10.1038/nature16961
102. Skinner B (1971) Beyond freedom and dignity. Bantam/Vintage, New York
103. Stawarczyk D, Bezdek MA, Zacks JM (2019) Event representations and predictive processing: the role of the midline default network core. Topics Cognitive Sci. https://doi.org/10.1111/tops.12502
104. Sugita Y, Tani J, Butz MV (2011) Simultaneously emerging Braitenberg codes and compositionality. Adapt Behav 19:295–316. https://doi.org/10.1111/tops.12502
105. Sutton R (2019) The bitter lesson (2019). https://doi.org/10.1111/tops.12502
106. Sutton RS, Barto AG (2018) Reinforcement learning: an introduction, second, edition edn. MIT Press, Cambridge
107. Sutton RS, Precup D, Singh S (1999) Between MDPs and semi-MDPs: a framework for temporal abstraction in reinforcement learning. Artif Intell 112:181–211
108. Szab ZG (2020th) Compositionality. In: Zalta EN (ed) The Stanford encyclopedia of philosophy, fall, 2020th edn. Stanford University, Metaphysics Research Lab, Stanford
109. Tani J (2017) Exploring robotic minds. Oxford University Press, Oxford
110. Tolman EC (1932) Purposive behavior in animals and men. Appleton, New York
111. Turing AM (1950) Computing machinery and intelligence. Mind 59:433–460
112. Ünal E, Ji Y, Papafragou A (2019) From event representation to linguistic meaning. Topics Cognitive Sci. https://doi.org/10.1111/tops.12502
113. Vinyals O, Babuschkin I, Chung J, Mathieu M, Jaderberg M, Czarnecki W, Dudzik A, Huang A, Georgiev P, Powell R, Ewalds T, Horgan D, Kroiss M, Danihelka I, Agapiou J, Oh J, Dalibard V, Choi D, Sifre L, Sulsky Y, Vezhnevets S, Molloy J, Cai T, Budden D, Paine T, Gulcehre C, Wang Z, Pfaff T, Pohlen T, Yogatama D, Cohen J, McKinney K, Smith O, Schaul T, Lillicrap T, Apps C, Kavukcuoglu K, Hassabis D, Silver D (2019) Alphastar: Mastering the real-time strategy game StarCraft II. https://doi.org/10.1007/s13218-015-0361-4
114. Watson J (1924) Behaviorism. Norton, New York
115. Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. IEEE Trans Evolut Comput 1(1):67–82
116. Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, Krikun M, Cao Y, Gao Q, Macherey K, Klingner J, Shah A, Johnson M, Liu X, Kaiser Ł, Gouws S, Kato Y, Kudo T, Kazawa H, Stevens K, Kurian G, Patil N, Wang W, Young C, Smith J, Riesa J, Rudnick A, Vinyals O, Corrado G, Hughes M, Dean J (2016) Google's neural machine translation system: bridging the gap between human and machine translation. ArXiv:1609.08144
117. Zacks JM, Speer NK, Swallow KM, Braver TS, Reynolds JR (2007) Event perception: a mind-brain perspective. Psychol Bull 133(2):273–293. https://doi.org/10.1007/s13218-015-0361-4
118. Zacks JM, Tversky B (2001) Event structure in perception and conception. Psychol Bull 127(1):3–21. https://doi.org/10.1007/s13218-015-0361-4
119. Zhu Y, Gao T, Fan L, Huang S, Edmonds M, Liu H, Gao F, Zhang C, Qi S, Wu YN, Tenenbaum JB, Zhu SC (2020) Dark, beyond deep: a paradigm shift to cognitive AI with humanlike common sense. Engineering 6(3):310–345. https://doi.org/10.1007/s13218-015-0361-4