RIVISTA INTERNAZIONALE DI FILOSOFIA E PSICOLOGIA
DOI: 10.4453/rifp.2023.0006
ISSN 2039-4667; E-ISSN 2239-2629
Vol. 14 (2023), n. 1-2, pp. 74-87

RICERCHE

From unified to specific theories of cognition

Frank van der Velde[(α)]

Ricevuto: 28 marzo 2022; accettato: 24 marzo 2023

Abstract This article discusses the unity of cognitive science that seemed to emerge in the 1950s, based on the computational view of cognition. This unity would entail that there is a single set of mechanisms (i.e. algorithms) for all cognitive behavior, in particular at the level of productive human cognition as exemplified in language and reasoning. In turn, this would imply that theories in psychology, and cognitive science in general, would consist of algorithms based on symbol manipulation as found in digital computing. However, a number of developments in recent decades cast doubt on this unity of cognitive science. Also, there are fundamental problems with the claim that cognitive theories are just algorithms. This article discusses some of these problems and suggests that, instead of unified theories of cognition, specific mechanisms for cognitive behavior in specific cognitive domains could be needed, with architectures that are tailor-made for specific forms of implementation. A sketch of such an architecture for language is presented, based on modifiable connection paths in small-world like network structures.
KEYWORDS: Connection Paths; Control of Activation; Small-world Networks; Symbol Manipulation; Unity of Cognition

Riassunto *Dalle teorie unificate della cognizione a quelle specifiche* – Questo articolo discute l'unità della scienza cognitiva che sembrava emergere negli Anni '50 e che era basata su una concezione computazionale della cognizione. Questa unità prevedeva l'esistenza di un singolo insieme di meccanismi (algoritmi) per tutti i comportamenti cognitivi, in particolare al livello della cognizione umana produttiva come, per esempio, linguaggio e ragionamento. A sua volta ciò implicava che le teorie psicologiche e, più in generale della scienza cognitiva, prevedessero algoritmi basati sulla manipolazione di simboli come nella computazione digitale. E, tuttavia, diversi sviluppi degli ultimi decenni hanno messo in dubbio questa unità della scienza cognitiva. Affermare che le teorie cognitive sarebbero solo algoritmi presenta problemi di fondo. Questo articolo discute alcuni di questi problemi, suggerendo che, invece di teorie della cognizione unificata, si potrebbe aver bisogno di meccanismi specifici per il comportamento cognitivo in specifici domini cognitivi, con architetture ritagliate per specifiche forme di implementazione. Questo articolo presenta uno schizzo di una simile architettura per il linguaggio, basata su vie di connessione modificabili in piccoli mondi come le strutture di reti.
PAROLE CHIAVE: Vie di connessione; Controllo dell'attivazione; Reti di piccoli mondi; Manipolazione di simboli; Unità della cognizione

[(α)]Cognition, Data, and Education, BMS, University of Twente, Drienerlolaan, 5 – 7522NB Enschede (NL)

E-mail: f.vandervelde@utwente.nl; veldefvander@outlook.com (✉)

1 Introduction

THE AIM OF THIS SPECIAL issue is to address the future of cognitive science(s). The formulation of this aim seems to imply the question of whether we could speak of future of cognitive science, or should speak of the future of cognitive sciences. I will address this issue in particular by focusing on the question of whether a theoretical foundation could be formulated that would account for the unity of cognitive science (both human-level cognition and artificial cognition).

Such a foundation seemed to have emerged in the 1950s, in a period that has been regarded as the beginning of cognitive science.[1] The reason for this is that, at that time, at least three developments came together: the shift from behaviorism to cognitive psychology, the start of artificial intelligence (AI) and the emergence of modern linguistics.

Behaviorism itself, as it arose in the first half of the 20th century, could perhaps be seen as a unified account of the basis of behavior (and in the work of Hull even of human and artificial cognition).[2] Yet, as formulated by Watson,[3] behaviorism originally aimed to explain human behavior based on learning only, specifically conditioning. So, if all human (and animal) behavior resulted from learning, this could imply that humans and animals are adapted to their environments, as these would be the sources of all their learned experiences. This leaves open the possibility that the mechanisms of cognition developed in this way are tailor-made for the specific link between the organism and its environment. In other words, there could be a close link between the specific cognitive architectures involved, the specific way they are implemented, and the cognitive domains they operate in. I will return to this possibility later on.

The emergence of the computational account of cognitive psychology and AI changed this view, at least for some time (e.g., up to the re-emergence of connectionism in the 1980s). A clear example is given by Newell's aim of «unified theories of cognition», which are all based on «a single set of mechanisms for all cognitive behavior».[4] In Newell's view, this single set of mechanisms is based on symbol manipulation as found in digital computing (e.g., in the von Neumann architecture). In turn, this would be needed to solve the problem of "controlled distal access" (or "logistics of access"),[5] which in Newell's view is required for any productive cognitive system, as I will discuss in section 5.1.

Another example of the computational view as a unified account of cognition is given by the critique on the re-emergence of connectionism in the 1980s by Fodor and Pylyshyn.[6] In their critique, they formulated three main features of human-level cognition, given by *productivity, compositionally* and *systematicity*, which in their view were not found in connectionist systems. Instead, they would require symbol manipulation, implemented in digital computational architectures.

The views of Newell[7] and Fodor and Pylyshyn[8] are closely related. In fact, the features *productivity, compositionally* and *systematicity* each require a computational architecture that possesses logistics of access as analyzed by Newell. Furthermore, these features also play a key role in the current debate on whether deep learning, as given in models like GPT-3, can provide human-level cognition.[9]

Therefore, I will rely in particular on Newell[10] and Fodor and Pylyshyn[11] as the basis for the view that computational architectures provide a unified theory of cognition. Here, however, I will only address a few issues related to this view. A discussion of other aspects related to this view can be found in van der Velde.[12]

Recent developments seem to cast doubt on whether unified theories of cognition as intended by Newell[13] would be possible, which also raises the question of what this would mean for the development of theories of human cognition. I will argue that the aim for unified theories of cognition may be out of reach. Instead, it would perhaps be better to look for specific cognitive architectures, implemented in specific ways and acting within specific cognitive domains. However, although this implies taking the cognitive domains and forms of implementation into account, it does not imply a simple return to behaviorism as we know it. To see why, I will start by briefly describing the transition from behaviorism to cognitive science.

2 From behaviorism to cognitive science

Around 1950, or even earlier, it became clear that classical behaviorism failed to explain human behavior in terms of conditioning only. The first mechanism explored was classical conditioning, based on an already existing (inborn) coupling between a stimulus and a response (reflex). This inborn response is then associated with a new stimulus, as illustrated with the famous study of Pavlov[14] on salivation.

Although classical conditioning undoubtedly works, it is problematic as the basis of all human behavior. Either because we would have to assume that all human behavior derives from just a few existing (inborn) responses, which makes the variability of human behavior difficult to explain. Or, we would have to assume a wide range of already existing, hence inborn, responses as the basis of human behavior, even for language and reasoning. This would be a very problematic assumption for a theory that boasted on explaining all of human behavior on the basis of learning alone, instead of inheritance. In the words of Watson: «we draw the conclusion that there is no such thing as an inheritance of capacity, talent, temperament, mental constitution, and characteristics».[15] This claim is

essentially meaningless if all our behavior derives from a large set of inborn responses.

However, the problem with classical conditioning seemed to be solved with the development of operant conditioning by Skinner. Instead of starting with an existing unconditioned response (behavior), operant conditioning could modify any form of behavior using reinforcement (e.g., reward). So if, say, a monkey accidentally pulled a lever upon which a reward emerged, the monkey would pull that lever more often.

Operant conditioning undoubtedly works as well. Because it can be used for any kind of behavior, it would seem it has solved the problem of classical conditioning. That is, on the assumption that the original behavior it starts with is just "random", inherited forms of behavior need not be assumed. The random behavior could then be modified into more purposeful behavior using reinforcement.

However, the notion of "random" behavior as the basis for learning is very unclear. In the case of a monkey initially pulling a lever, it is clearly not the case that the animal moves its arms and legs randomly until one of them accidentally pulls the lever. Instead, it would look selectively at the lever first and then pull it. Hence, the initial behavior should be seen more as explorative than as random, which again begs the question of where it comes from.

Other examples of explorative behavior were found in behaviorist experiments as well.[16] For example, a rat left on its own in a maze would be able to find the shortest path even before it was trained (rewarded) to do so. To deal with this problem, behaviorists introduced the concept of the "drive", like a "curiosity" drive that would produce the exploratory behavior of an animal. Drives such as curiosity are then rewarded by the behavior of the animal. This would eliminate the need for an external reward to account for learning, but would maintain the idea that all behavior results from learning based on some reward.

But again, drives have to be assumed to exist beforehand, which diminishes the importance of learning in explaining behavior. Also, the sheer amount of different drives needed resulted in an incoherent view about their nature and their relations. In the 1950s, all of this resulted in the shift from behaviorism to cognitive psychology, which emphasized that behavior results from the processing of information, instead of just learning associations.

However, it is remarkable to see that a similar criticism of the behavioristic approach was already formulated much earlier.[17] In the 1910s, Köhler studied how chimpanzees solved so-called "detour" problems. For example, a banana would be visible but just out of reach. However, by making use of material available (e.g., wooden boxes to be used for climbing) the chimpanzees were able to solve the problem, in that they could get access to the fruit. Although these studies were conducted before the development of operant conditioning, Köhler already noticed the problem of relying on initial random behavior to get learning started. In his words:

> In the description of these experiments it should have been apparent that the chief essential is lacking for an explanation by chance actions, that is to say, the means by which the solution is composed out of chance parts is not apparent. Certainly it is not a characteristic of the chimpanzee, when he is brought into an experimental situation, that he should make chance movements out of which, among other things, a non-genuine solution could arise. Very seldom is a chimpanzee seen to attempt any action that would have to be considered accidental in relation to the situation [...] all distinguishable stages of his behavior [...] tend to appear as complete attempts at solutions, of which none appears as the result of accidentally arranged parts. [...] Never, in real and convincing cases, does the solution merge from the disorder of blind impulses. The action is smooth and continuous and can be resolved into parts only *by the abstract thinking* of the observer. In *reality* the parts do *not* appear independently. Thus [...] our theory cannot permit the supposition that [...] the solutions that came as wholes could possibly have arisen from mere chance.[18]

This quote and its timing raise the question of why it took until the 1950s before the conclusion could be reached that human (and animal) behavior cannot be (fully) described by means of learning based on conditioning. One reason could be that behaviorism simply had to run its course, before its limitations became more apparent and convincing.

Another reason might be found in philosophy, in particular the philosophy of science. Positivism and later logical positivism were dominant in the first half of the 19th century, and it is clear that Watson was strongly influenced by it.[19] Although Popper had already criticized aspects of logical positivism in the 1930s, it was not until the 1950s that it lost its dominant position in the philosophy of science, based on the work of, e.g., Hanson[20] and Kuhn.[21] It is an interesting question (but beyond the scope of this article) to see if there is indeed a relation between the rise of post-positivism and cognitive psychology (insights from Gestalt psychology were certainly used in post-positivism).

Yet another and perhaps decisive reason could have been the development of the computer, not only as an abstract model of information processing, as with the Turing machine, but also as a practical tool. This gave the idea of what information processing could be and how it could be

developed and tested, as exemplified with the start of AI. Initially, cognitive psychology and AI developed along similar lines, based on the idea that cognition derives from computational processes in the form of symbol manipulation.[22]

The emergence of modern linguistics strengthened the notion that cognitive processing is based on symbol manipulation. In 1957 Skinner published a book on verbal behavior, in which he argued that we produce and understand a sentence based on learned associations between words.[23] In the same year Chomsky published a book in which he argued that a sentence has a syntactic structure, which cannot be understood as just an association between words.[24] The notion of syntactic structures and the program (grammar) needed to produce and analyse them fitted very well with the computational approach in cognitive psychology and AI that developed around the same time.

3 Competing approaches on the nature of cognition

In the 1950s it seemed that there is just a single cognitive science, dealing with cognition as a form of symbol manipulation (both for human cognition and artificial intelligence). However, since then competing approaches on the nature of cognition, such as connectionism and dynamical approaches, emerged. On its own that would not indicate the end of a single science of cognition. In Kuhn's terms, it could indicate instead that this science has not found its foundational paradigm yet.

However, more recent developments in AI do seem to cast doubt on the unity of cognitive science, as illustrated with Alpha Go and Alpha Go Zero.[25] The first is an AI program that learned to play the game Go and succeeded in beating the world champion. Later, a similar program was developed for Chess, with a similar result. However, Alpha Go itself was defeated by Alpha Go Zero (100 to 0, both in Go and Chess). A remarkable difference between these two programs resides in the way they were trained. The Alpha Go program was first trained by using knowledge that had been acquired (by humans) on how to play Go or Chess. Then, the program was developed further by playing against itself, using forms of machine (reinforcement) learning. In contrast, the only forms of knowledge used to train Alpha Go Zero were the rules of the game Go or Chess. Then, the learning procedure based on playing against itself was used to develop the program further.

The significant defeat of the Alpha Go program by Alpha Go Zero raises an important question about the human knowledge on how to play Go or Chess, used with the first program. One could assume that, during the ages, humans would have acquired a lot of knowledge on how to play, e.g., Chess, such as the best ways to start the game or how to respond to the opponent in certain situations. So, it would seem that using that knowledge would be a benefit for an AI program. It would, so to say, have a kick-start with this knowledge and then could learn further. It would certainly be a benefit over a program that was not given this kick-start, but had to find out everything for itself. Or so it would seem.

But the results show otherwise. The knowledge on Chess (or Go) used as kick-start apparently hindered the program it is development. It seems as if it was burdened by it and had to unlearn it before it could learn to play Chess in the proper way. But a program not burdened by human knowledge on Chess would develop further and would have the upper hand. The minimal conclusion from this is that, apparently, we do not understand the game of Chess; at least not in the way it could be understood. This raises the question of what that knowledge would be. Clearly, it is engraved in all of the relations etc. learned by the Alpha Go Zero program. But how are we to understand it, even if we could analyse all of these relations?

The example of Alpha Go Zero suggests that our knowledge (cognition) is different from the cognition that could be acquired with certain forms of machine learning. This raises the question of whether the reverse could also be true, and if so, what that would entail about the unity of cognitive science. An example is found in the language behavior program GPT-3.[26] This program is based on a neural network and trained on a huge amount of sentences (more than humans see in their lifetime). It can respond to questions or situations by producing answers in fluent English, which suggests that it has mastered the language.

However, Marcus and Davis[27] analyzed the behavior of GPT-3 with a number of scenarios. In each case, GPT-3 was given a brief description of a situation, to which it responded. Although this research was not yet intended as a systematic investigation of GPT-3's abilities on language completion or reasoning, a few observations do stand out from the replies given.

Firstly, GPT-3 will produce a response to the scenario given whether or not that response "makes sense" (e.g., is actually or even remotely related to the given scenario). Secondly, even when the response makes no sense in any meaningful way, it is not entirely random. For example, one scenario concerned the use of a cigarette to stir a drink (when a spoon is not available). GPT-3 replied by telling a story about crematoria. This story had nothing to do with the issue at hand, which suggests that it did not understand what the issue was about.

It is, of course, possible not to understand an issue. But, in general, a cognitive agent would (should) be able to reflect on that and acknowledge that it does not understand. From the responses given by GPT-3 on this and other scenarios it seems that it

does not have that form of reflective knowledge. That is, it cannot make a distinction between what it knows and what it does not. Instead, it will just give an answer, apparently based on direct or indirect associations it has learned in training. For example, indirect associative links between cigarettes and crematoria would certainly be present in the learning material for GPT-3, such as the use of fire, the production of ashes, or the strong association with death.

The inability to reflect on what you know or not and giving an answer in all cases regardless of whether it makes sense constitutes a real problem for language understanding, and reasoning (cognition) in general. This could indeed be a difference between learning language and learning to play Chess or any other game. In the latter case, there are real restrictions on what you can do, as given by the rules of the game, the space that the game is confined to, or other constraints on the moves you can make. So even it, say, a Chess program would produce a move like Knight to H9, it would be restrained from performing it because it is not possible.

In language, and cognition in general, such restrictions are far less clear. The examples with GPT-3 indicate that, apparently, these restrictions cannot come from learning sentences only (as noted, GPT-3 is already trained on more sentences than a human will encounter in a lifetime). They would also have to come from learning about the world in a more direct manner. Moreover, humans learn in a different way. Children develop their ability for language in an incremental manner, learning brief sentences and simple scenarios first. Incremental learning is very difficult for neural networks as used in machine learning, because it intervenes with the statistical analysis of the data these networks develop in learning.[28]

So, there is a possibility that human cognition and cognition acquired with certain forms of machine learning are distinctly different. One could argue here that this conclusion is premature, because further developments with programs like GPT-3 might eventually produce programs that have learned to understand the world in the way that humans do.

But this argument misses the point. Even if it were possible to develop AI systems at the level of human cognition, it is apparently also possible to develop AI systems that are significantly different. Of course, one could still subsume all of this under the same heading of "cognitive science", but a key issue for that science would then be to understand why, apparently, different forms of cognition are possible, and what that would entail for the general notion of "cognition".

A first attempt to do this is to have a closer look at the unity of cognitive science that seemed to emerge in the 1950s. As outlined above, the development of the computer played an important role in this, as is also clear from the view of Fodor and Pylyshyn[29] in their well-known analysis of connectionism. They argued that this constituted a return to behaviorism. On that note, Watson's[30] version of behaviorism was based on a positivist's view of psychology as a science. So, he was not interested in making models. However, the behaviorist Hull was very much interested in developing models on how behavior could be produced.[31] His models were hand designed and consisted of long chains of stimulus-response associations (reflexes). But in their appearance they are not so different from, say, feedforward neural networks as used in connectionism.

According to Fodor and Pylyshyn, the architecture underlying cognition must be a computational architecture as found in digital computing, such as the Turing machine or the Von Neumann architecture. The reason is that these architectures provide the means to process symbolic structures in a rule-based manner. In turn, this is needed to provide the main features of human-level cognition, given by the related features of productivity, compositionality and systematicity.[32]

An example is given by our ability to understand arbitrary "who does what to whom" relations in arbitrary sentences.[33] Systematicity, for example, implies that if you understand that *Sue* is the agent in *Sue eats pizza*, you cannot but understand that *pizza* is the agent in *pizza eats Sue*, even though that is semantically odd. Indeed, we know that *pizza eats Sue* is odd precisely because we identify *pizza* as the agent and *Sue* as the object (theme) of *eat*.[34]

So, it is no surprise that these features concur with Chomsky's[35] view on the unlimited productivity of language, which in the view of Fodor and Pylyshyn would be achievable only with computational architectures. Here, I want to focus on one aspect of computation as referred to by Fodor and Pylyshyn. It concerns the role of implementation in theories of cognition, discussed in the next section.[36]

4 The role of implementation in cognition

The topic can be introduced by a quote from Fodor and Pylyshyn on whether it would be useful to know how the architectures they refer to are actually implemented in the brain. Their response is:

> The answer [...] has always been that the *implementation*, and all properties associated with the particular realization of the algorithm that the theorist happens to us in a particular case, is irrelevant to the psychological theory; only the algorithm and the representations on which it operates are intended as a psychological hypothesis.[37]

An algorithm is indeed independent of the way it is implemented. This follows from the theory of computable or recursive functions.[38] These are

functions for which an "effective procedure" can be found to compute the function. Recursive function theory shows that a formal definition of what this means cannot be given. But it also shows that all procedures developed thus far are equivalent with (or lesser than) the Turing machine. Turing[39] developed the Turing machine to give an answer on the question of what an effective procure could be. In short, it is a program that can be executed on the Turing machine.

Fodor and Pylyshyn do indeed see the Turing machine, and related architectures such as the Von Neumann architecture, as the basis for cognitive architectures. In this way, one can describe what the aim of a unified cognitive science would be. In line with Newell,[40] it would consist of developing and studying the "algorithms and the representations on which they operate", as asserted in the quote above. Hence, this approach equates cognitive theories and models with algorithms, and thus equates any cognitive task with a computable function.

The fundamental problem with this approach is that mathematical functions, of which computable functions are a subset, are inherently static. This is clear from the general definition of a mathematical function, which describes a function as a relation between two sets, the domain (input) and the range (output). The function is characterized by the way it assigns an element from the range to an element of the domain. For a subset of functions, a description of this relation can be given. For example, the numerical function $f(x) = 2x$ assigns the output $2x$ to the input x.

Another subset of mathematical functions is the set of computable functions. For these functions, so-called "effective procedures" or algorithms can be given that produce an output given an input.[41] The function $f(x) = 2x$ belongs to this subset as well, because there are algorithms on the Turing machine and other computers that produce the value of $2x$ for the input x (assuming sufficient memory is available). It is clear now why implementation plays no role here. For example, the output of $f(x) = 2x$ for $x = 3$ is given by 6, because that follows from the function description. The role of the algorithm is only to produce the output 6 for the input 3, otherwise it would in fact compute a different function. The implementation of an algorithm could affect, for example, the time it takes to compute the output, but not the output itself. It makes no sense to say that "today the output of $f(3)$ is 5, because there is no time to compute further".

But what about cognition? Picture a hominid living on the planes in Africa, who is confronted with an animal. Let's say that the choice here is between a lion or a deer. At face value, this looks like a functional problem, and indeed the problem has a clear functional aspect. But it is also clear that the behavior of the hominid, and indeed its survival, critically depends on the time in which an answer is "computed". This aspect is not included in the definition of a mathematical function. And yet, it is crucial for cognition, because the most fundamental aspect of cognition is to generate the behavior that enhances survival.[42]

Hence, time is an important factor in cognition but it is (by definition) not included in computation theory. This shows that Fodor and Pylyshyn[43] are wrong: "psychological hypotheses" cannot be only "algorithms and the representations on which they operate". However, as the choice between a lion or deer shows, there are "functional" aspects to the production of behavior. These functional aspects can be integrated with a dynamical (time) constrained description in terms of dynamical systems as functional "flows".[44]

It is important to understand that this problem cannot be solved by including time as a factor in an algorithm. For example, a computer program that generates a weather prediction will include time as a factor, because e.g. you want to know when the storm arrives and how long it will last. However, this program can be executed on two computers with the same computational precision but one faster than the other. The result will be that both computers generate the same weather prediction, because they run the same algorithm. But the execution times of the program will be different. This shows that the execution time of an algorithm is not a part of the algorithm itself. Yet, it is a part of many cognitive tasks, as illustrated above. So, the processing underlying these tasks is not only depended on an algorithm (i.e. an effective procedure for a computational function).

The main conclusion here is that the computational characterization of cognition that emerged in the 1950 is at best incomplete. This casts doubt on the unity of cognitive science, as it seemed to emerge in that period. Cognition is not just computational, or better, functional (in the mathematical sense of the word). It also depends on the satisfaction of constraints such a speed of processing needed for survival and potentially other constraints as well. These constraints and their relations could be different for different cognitive domains. In particular, certain forms of implementation could be selective for certain forms of information processing.

A glimpse of that could be found in the game of Jeopardy that IBM's Watson played against the two best human players at that time.[45] IBM's Watson won the competition. But there were a number of problems that the humans solved better. Not because IBM's Watson did not know the answer. It did, but it was too slow in these cases. The (likely) reason of why IBM's Watson was too slow is that it treated each problem in the same (analytical) manner. For the humans, however, some problems were easier to solve, presumably because

of the stronger activation of associations in their memory in these cases. In turn, this indicates or suggests that human memory is selective. This, again, would be an effect of implementation, as given by the way the brain learns, stores and retrieves information, specified for the environments or domains in which it operates.

Viewed in this way, the competing approaches on the nature of cognition could be more than just an indication that cognitive science has not found its foundational paradigm yet. It could also mean that they target different domains, based on different architectures that are tailored to the way they are implemented. So, those that succeed in winning games like Go and Chess in a way that is perhaps beyond our understanding would not necessarily be the best suited for language. At least not concerning the interactions of language with the environment, as given by the ability to deal with multiple constraints in a limited amount of time, as well as the need for incremental learning.

At this moment, these are just mere suggestions, for which future developments will show to what extent they are true. But, at least, they put a focus on closer interactions between the domains of cognitive processes, the underlying architectures, and the way the architectures are implemented (and are influenced or determined by their implementation). In case of human cognition, this would require a more profound understanding of how cognitive processing relates to the structure and dynamics of the brain. Much more indeed than anticipated or advocated by Fodor and Pylyshyn.[46] The next section discusses this relation in somewhat more detail.

5 Cognition implemented in the brain

In his *Principles of psychology*, William James stated the following assertion about the relation between cognition and brain (or psychology and neuroscience, if you will):

> For the entire nervous system *is* nothing but a system of paths between a sensory *terminus a quo* and a muscular, glandular, or other *terminus ad quem*.[47]

This quote relates to the central thesis of modern neuroscience, as initiated by Cajal, that neurons form connection paths in the brain.[48] But, in my view, it also reflects a deep insight into what cognition is about and how we should aim to understand it. When push comes to shove, the aim of cognition is to provide the organism with better changes of survival, which is indeed reflected in the ability to act in response to the stimulation received from the environment.

This entails a behavioristic component in cognitive science. However, to be clear, it does not indicate a return to classical behaviorism, just based on conditioned reflexes and drives. As noted in section 2, this was motivated by a positivist's view on the way science (psychology) should operate, which is not implied in James' quote. This quote does not imply either that models of how the brain produces behavior should be reflexive in their nature. As analyzed by Amsel and Rashotte,[49] this was in fact the main problem of Hull's behavioristic models. Instead, cognition «intervenes in the sensorimotor loop by means of which the creature interacts with its physical and social environments».[50]

In the case of human behavior, all aspects one would attribute to higher-level cognition play a role in this intervention. Fodor and Pylyshyn[51] do in fact make a strong case for productivity, compositionality and systematicity as important features of human-level cognition, as exemplified in language. Incremental learning should be added here as well, also because it is closely related to these features.

Incremental learning entails that a child will learn language starting with small sentences, and is capable of gradually integrating already learned material with newly acquired knowledge. This is very difficult for models like GPT-3, which have to be retrained extensively when they aim to acquire new information. In machine learning with neural networks, re-learning already learned material is needed to prevent undoing the learned knowledge by newly learned material (sometimes referred to as "catastrophic interference"). This behavior is not just an accident but derives directly from the way these networks learn.[52] An example is found with GPT-3. At some point a mistake was detected in the set-up of the training data. However, to address this afterwards would have required substantial relearning. Due to the costs involved it was decided not to do this.[53]

However, the architecture that provides these features of cognition would have to be implemented in terms of the structure and dynamics of neural processing in the brain. This requirement, in my view, rules out symbolic architectures that are implemented in a neural manner.[54] But, as briefly illustrated in the next section, it could be achieved by an architecture that would provide the ability to control connection paths between perception and action, that is, control of connection paths that "intervene in the sensorimotor loop". Features such as productivity and compositionality could be implemented if the architecture has a connection structure that resembles a small-world like network.[55] In particular, because the logistics of access needed for these features could be implemented in this way.[56]

5.1 Combinatorial productivity in a small-world like network structure

A key feature of human language is "combinato-

rial productivity", which concerns the virtually unlimited ability to combine words in arbitrary sentence structures. The ability to combine sound patterns is found in animal communication as well,[57] which would make the difference in combinatorial productivity between humans and animals merely quantitative. But this difference in quantity is in fact so huge that it becomes a different quality in its own right. The fact that human cognition is singled out by its combinatorial productivity is illustrated in, for example, movies and cartoons. Young children have no difficulty in relating to a character like an artificial sponge living in a pineapple at the bottom the sea, even though that is not derived from their direct experience.

Here, incremental learning and combinatorial productivity come together. Children will have learned what an "agent" is and what it means to live in a house. Effortlessly, they can then recombine these roles and relations even with fantasy characters and environments. For hominids, combinatorial productivity would have been a key ability for their survival, and it is a key ability in the environments we live in, much more so than, say, the ability to play Go or Chess.

Behaviorism, even in the way of Hull, would fail on this. Combinatorial productivity cannot be achieved on the basis of associations alone. For one, it cannot provide associations for relations between agents, or between agents and actions, never seen before. These include, for example, potential relations between agents that are the reverse of relations between those agents that have been learned. So, if a child has learned that a cat could chase a bird, it could also understand what happens if a bird would chase a cat, even though it has not yet seen that, or even if it would never actually occur. Again, this is an ability that makers of, say, movies or cartoons are relying on. It is also an ability that relates to the notions of systematicity and compositionality, as analysed by Fodor and Pylyshyn.[58]

As noted in the introduction, combinatorial productivity in computational terms requires architectures that possess the ability to achieve controlled distal access to information and integrate it in processing,[59] which I refer to here as logistics of access.[60] As analyzed by Newell,[61] the need for this ability derives from the fact that, in physical terms, the amount of information that can be stored at a local site is limited. So, with more information required, the architecture needs to have distal access to that information.

An example is given by the internet. The information we can store on a local computer is limited, but we can obtain more information by means of distal access to other computers, and then download that information to influence processing on the local computer.

Another example is found in language. With a lexicon of 60.000 words or more,[62] word information will be stored at different sites in the architecture (as it is in the brain).[63] So, distal access to these sites is needed to integrate these words in a sentence structure. This includes the ability to integrate newly learned words directly in sentence structures. As noted above, integrating newly learned information is hard for models such as GPT-3, which is a strong indication that these models do not possess logistics of access.[64]

Computational architectures such as the Turing machine and the Von Neumann architecture do possess logistics of access, which is of course the reason why these architectures seemed to be the basis for cognitive science as it emerged in the 1950s. And, as outlined by Newell,[65] they achieve logistics of access by using symbols, e.g., to retrieve information and copy it so that in can affect local processing (as in the internet example given above). As a result, cognitive processing would have to consist of forms of symbol manipulation in this view.

But, as argued above, there are serious issues with these architectures from a cognitive perspective. There are also serious issues with them from a neural perspective. For example, Kreite and colleague[66] proposed that neural codes standing for words are stored in dedicated registers to represent a sentence. So, there would be registers for verbs and for nouns as agents or themes of a verb. This would allow the representation of a sentence like *Bob ate steak*. Indeed, this is how cognitive architectures based on symbol manipulation represent sentences.[67] But the idea that words are represented in the brain as neural codes that could be (and would have to be) transported to dedicated registers (e.g., using a data bus as in the Von Neumann architecture) to represent a sentence does not match with what is known about the way conceptual information is stored in the brain.

Hebb already suggested that conceptual information would be represented in the brain as interconnected structures he referred to as neural assemblies.[68] These could and would be distributed over the brain and would develop gradually, as more information about a concept is learned over time. More recent investigations of the structure of semantic memory corroborate this view.[69]

A representation of a concept as a Hebbian assembly precludes its use as a neural code that could be stored in registers. Instead, it will always remain "*in situ*"[70] even when it is used as a word in a sentence structure. A sentence structure is then a connection path between these in situ word (concept) representations.

However, the logistics of access needed for combinatorial productivity imposes major demands on the ability to form connection paths in the architecture. For example, connection paths must be possible between all in situ assemblies of nouns and all in situ assemblies of verbs, in which any noun could be any of the arguments of any

verb (agent, theme or recipient, depending on the verb). This includes combinations never seen before, such as *bird chases cat*.

As an additional problem, it is unlikely that new connections between neurons or populations are or can be created on the fly. Of course, modifications of existing connections (in particular in the hippocampus and surrounding areas)[71] can occur in short time windows, as in the time window for sentence processing. But growing new connections that would interconnect novel combinations of in situ concept assemblies during verbal communication seems unfeasible.

Given that new connections are not created on the fly in verbal communication, the logistics of access needed for combinatorial productivity has to be achieved with a "fixed" connection structure. This demand holds for any model of neural language represen.tation or processing. For example, Kriete and colleagues' model[72] has to account for the fact that the neural code of the word *steak*, or indeed of any other noun, can be stored in each of the registers for nouns. Given that new connections are not created in that process, the model has to assume a fixed connection structure in the brain by which this can be achieved. An example would be a connection structure similar to the Von Neumann architecture of the digital computer. In this architecture, the content from any arbitrary register can be transported (copied) to any other register by using the central data bus, controlled by the CPU.

However, the computing architecture in which sentence structures are connection paths interconnecting in situ word representations would be very different from the Von Neumann architecture with its data bus controlled by the CPU. Instead, the suggestion is that it could be achieved with a particular connection structure related to the notion of a "small world network".[73]

Here, only a brief outline of the motivation of this idea can be given, as illustrated in *Figure 1*.[74]

Panel (a) of *Figure 1* illustrates a connection structure in which every noun is connected to every verb as its agent. A similar connection structure would then have to exist for the theme and recipient arguments of verbs (as with the verb *give*). Two problems arise here. Firstly, the extensive amount of connections needed, given the number of nouns and verbs adults are familiar with (in the order of 10.000 each).[75] Secondly, it is very hard to see how new nouns and verbs could easily be added to the structure. Yet, we learn new nouns and verbs throughout our lives.

In panel (b), a (very!) basic solution is presented by the introduction of "agent nodes" (neurons, neural populations) that exist in between the nouns and verbs stored in the brain. These agent nodes reduce the overall connection structure needed to connect nouns as agents of verbs. In-

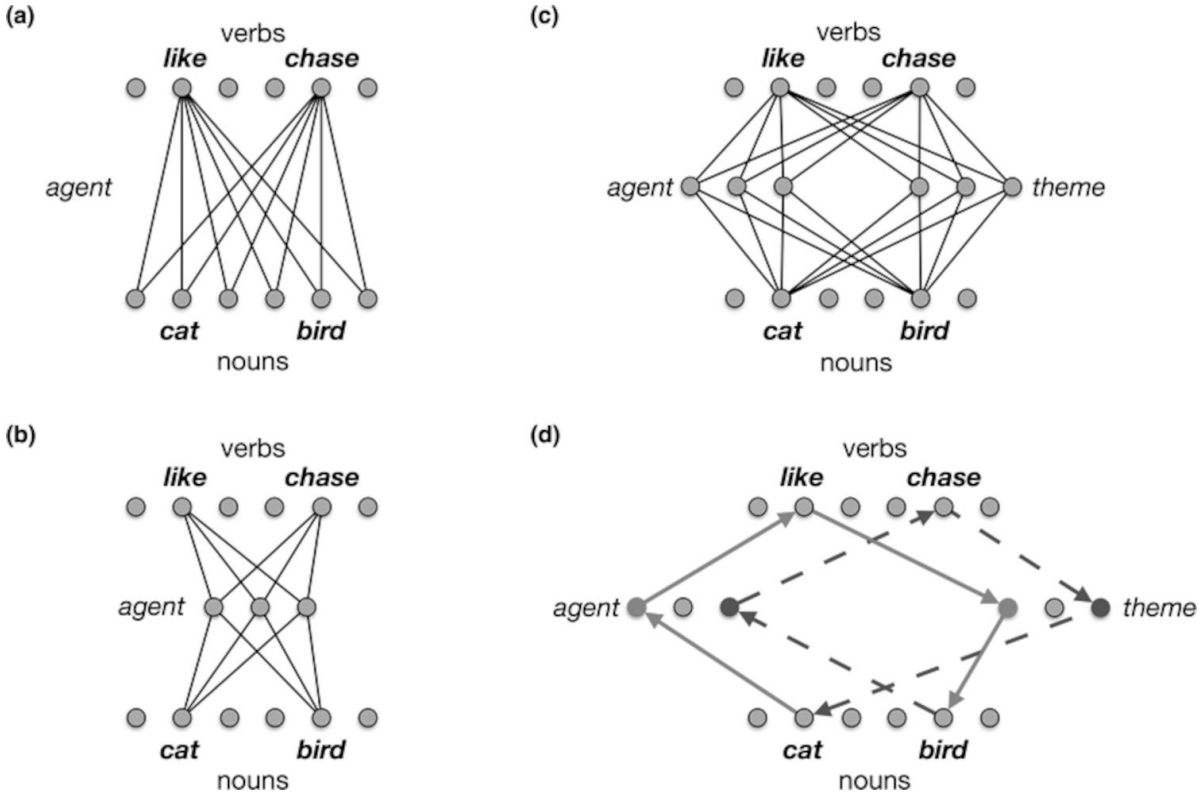

Figure 1. Illustration of connection structures between nouns and verbs. (a) Connections between all noun and verb representations. (b) Connections between nouns and verbs via agent nodes. (c) Combination of agent and theme nodes to connect nouns and verbs. (d) Connection paths for the sentences *Cat likes bird* (red solid) and *Bird chases cat* (blue dashed) in the connection structure of (c).

stead of every noun being connected to every verb, each noun is connected to a more limited set of agent nodes only, which in turn are connected to each verb.

The core of the solution illustrated in *Figure 1* is that the agent nodes operate as the "hubs" one finds in small-world networks. On the one hand, the hubs significantly reduce the number of connections needed, as illustrated by the difference between the panels (a) and (b) in this figure. And yet, on the other hand, the overall connectivity, in which every noun can be an agent of every verb as found in (a), is retained. In (c), the agent nodes are combined with theme nodes that operate on the same principle.

Panel (d) illustrates how connection paths could be established in this connection structure that represent sentences, such as *Cat likes bird* (red-solid) and *Bird chases cat* (blue-dash). The arrows indicate the order of activation in the connection paths that corresponds with hearing or speaking these two sentences. The figure also shows schematically that a word can occur concurrently in two different sentences, and in two different roles.

Of course, the simple schematic connection structure in *Figure 1* raises a number of deep issues. Van der Velde and de Kamps presented in 2006 a first attempt to solve them.[76] It shows that a reduction in connections can be achieved in this way, and that the introduction of new nouns (or verbs) is easier to account for. In terms of *Figure 1*, a new noun (verb) needs to be connected only to the limited set of agent nodes, instead of to all verbs (nouns). More recent developments of the underlying architecture show that arbitrary sentences in English can be represented and processed as connection paths in this way.[77]

The key aspect of the solution as outlined in *Figure 1* is that it provides the logistics of access needed for combinatorial productivity in a manner that is different from the Von Neumann architecture. In both cases, controlled distal access is needed to information outside a local site. This follows from the physical constraint of the amount of information that can be stored at a local site, and thus applies to any architecture that aims to achieve combinatorial productivity.

However, the Von Neumann architecture uses symbols to retrieve information from a distal site to affect local processing (e.g., as illustrated in the internet example given above). In contrast, in the solution outlined in *Figure 1* information is not stored with symbols. Instead, it is embedded in the network structures at any site in the architecture, as suggested by Hebb[78] and illustrated by Huth and colleagues.[79]

Distal access is then achieved by temporarily interconnecting distal and local information in a connection path in the architecture, so that information at both sites can be integrated. In this process, the information itself remains in situ at all times. The network structure that allows the creation of these connection paths it provided by the small world network structure as outlined in Figure 1. Processing in this architecture does not consist of forms of symbol manipulation (as there are no symbols in the architecture) but by means of controlling the creation and use of the connection paths in the architecture. In turn, these differences in processing could have an affect on behaviour, in particular under time constraints.

However, the fact that full combinatorial productivity, as in language, can be achieved in this way shows that this solution is viable alternative for both the von Neumann architecture, which achieves productivity with symbol manipulation, and forms of deep learning as in GPT-3, which as yet lack in full combinatorial productivity.[80]

Furthermore, simulations of the architecture show that sentence processing proceeds differently in architectures like this compared to symbolic processing. These differences emphasize the importance of implementation, even in the case of productive architectures that can represent and process arbitrary "who does what to whom" relations in arbitrary sentences, which is a core feature of human language.[81]

6 Conclusions

As Newell argued, the unity of cognitive science would be based on a "single set of mechanisms for all cognitive behavior",[82] consisting of computational processing as found in digital computing. Fodor and Pylyshyn[83] corroborated this view by equating theories in psychology (cognition) with algorithms based on symbol manipulation.

However, a number of arguments suggest that such a "single set of mechanisms" might not exist. First of all, the equation of theories in psychology or cognitive science with algorithms is simply wrong. It ignores effects of implementation, such as speed of processing. These are irrelevant for algorithms, but important and sometimes vital for cognition. Secondly, different approaches for specific cognitive domains have emerged in recent years, next to the architectures based on symbol manipulation. Some of these have obtained stunning successes, as in playing games like Go and Chess in ways that seem to exceed human understanding.

The combination of different approaches with the importance of implementation for cognition suggests a shift from unified to specific theories of cognition. Each of these would deal with a specific cognitive domain, with architectures that are tailor-made for specific forms of implementation. Of course, they could all still be subsumed under the label "cognitive science", but this would not fulfill the aim that Newell had with unified theories of cognition.

As noted, the importance of implementation and the constraints imposed by a cognitive domain (for example, the need for survival) entail a return of a behavioristic component in cognitive science. The role of reinforcement learning in machine learning, and indeed the emphasis on learning as the basis of cognition in this field underscores this return. However, it would be a mistake to discard fundamental features of cognition recognized after, and even before, the decline of behaviorism in the 1950s. Among these are the productivity, systematicity and compositionality of human level cognition, as well as the importance of incremental learning.

In the case of human cognition, insight in the role of implementation will be provided by the interaction between neuroscience and psychology. New kinds of architectures have to be developed to satisfy the constraints of both the structure and dynamics of the brain and the fundamental features of cognition and behavior. I illustrated this with a sketch of an architecture in which sentence structures consist of (temporal) connection paths interconnecting in situ word or concept representations, as found in human semantic memory.

Further developments of such architectures are needed. But their successes would show that it is possible to implement fundamental cognitive features like productivity, systematicity and compositionality without relying on forms of symbols manipulation. Hence, instead of relying on a "single set of mechanisms", they would be achieved with different mechanisms, based on activation control of connection paths in small-world network structures as provided by the structure of the brain.

Notes

[1] Examples are the Dartmouth Project in 1956, which is generally seen as the start of the field of Artificial Intelligence (cf. *The Dartmouth summer research project on artificial intelligence*, available at URL: https://home.dartmouth.edu/about/dartmouth-milestones, accessed 14 March 2023), and the review of Skinner's book on language by Chomsky, which contributed to the decline of behaviorism as the foundation of human-level cognition (e.g., cf. B.M. THORNE, T.B. HENLEY, *Connections in the history and systems of psychology*).

[2] Cf. A. AMSEL, M.E. RASHOTTE (eds.), *Mechanisms of adaptive behavior*.

[3] Cf. J.B. WATSON, *Psychology as the behaviorist views it*; J.B. WATSON, *Behaviorism*.

[4] A. NEWELL, *Unified theories of cognition*, p. 15.

[5] Cf. F. VAN DER VELDE, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*.

[6] Cf. J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture: A critical analysis*.

[7] Cf. A. NEWELL, *Unified theories of cognition*.

[8] Cf. J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture: A critical analysis*.

[9] Cf., e.g., J. BROWNING, Y. LE CUN, *What AI can tell us about intelligence*. In: «Noema», available at URL: https://www.noemamag.com/what-ai-can-tell-us-about-intelligence/, accessed 30 September 2022; G. MARCUS, *The next decade in AI: Four steps towards robust Artificial Intelligence*.

[10] Cf. A. NEWELL, *Unified theories of cognition*.

[11] J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture: A critical analysis*.

[12] Cf. F. VAN DER VELDE, *Towards a neural architecture of language*; F. VAN DER VELDE, *The neural blackboard theory of neuro-symbolic processing.*

[13] Cf. A. NEWELL, *Unified theories of cognition*.

[14] Cf. I. PAVLOV, *On conditioned reflexes* (1904).

[15] J.B. WATSON, *Behaviorism*, p. 94.

[16] Cf., e.g., D. HOTHERSALL, *History of psychology*.

[17] This example shows that, even when a particular view is dominant at a given moment, different views could also, and often do, exist at the time. Examples of this in case of the dominance of the computational view in the 1950s to 1970s would be Hebb (D.O. HEBB, *The organisation of behavior*) or Rosenblatt (F. ROSENBLATT, *The perceptron*).

[18] Cf. W. KÖHLER, *On the insight of apes* (1917), p. 571 – italics added.

[19] Cf. J.B. WATSON, *Psychology as the behaviorist views it.*

[20] Cf. N.R. HANSON, *Patterns of discovery*.

[21] Cf. T.S. KUHN, *The structure of scientific revolutions*.

[22] Cf., e.g., Z.W. PYLYSHYN, *Computation and cognition*; A. NEWELL, *Unified theories of cognition*.

[23] Cf. B.F. SKINNER, *Verbal behavior*.

[24] Cf. N. CHOMSKY, *Syntactic structures*.

[25] Cf. D. SILVER, J. SCHRITTWIESER, K. SIMONYAN, I. ANTONOGLOU, A. HUANG, A. GUEZ, T. HUBERT, L. BAKER, M. LAI, A. BOLTON, Y. CHEN, T. LILLICRAP, F. HUI, L. SIFRE, G. VAN DER DRIESSCHE, T. GRAEPEL, D. HASSABIS, *Mastering the game of Go without human knowledge*.

[26] Cf. T.B. BROWN, B. MANN, N. RYDER, M. SUBBIAH, I. KAPLAN, P. DHARIWAL, A. NEELAKANTAN, P. SHYAM, G. SASTRY, A. ASKELL, S. AGARWAL, A. HERBERT-VOSS, G. KRUEGER, T. HENIGHAN, R. CHILD, A. RAMESH, D.M. ZIEGLER, J. WU, C. WINTER, C. HESSE, M. CHEN, E. SIGLER, M. LITWIN, S. GRAY, B. CHESS, J. CLARK, C. BERNER, S. MCCANDLISH, A. RADFORD, I. SUTSKEVER, D. AMODEI, *Language models are few-shot learners*.

[27] Cf. G. MARCUS, E. DAVIS, *GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about*.

[28] Cf. F. VAN DER VELDE, *Computation and dissipative dynamical systems in neural networks for classification*.

[29] Cf. J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture: A critical analysis*.

[30] Cf. J.B. WATSON, *Psychology as the behaviorist views it.*

[31] Cf. A. AMSEL, M.E. RASHOTTE (eds.), *Mechanisms of adaptive behavior*.

[32] Cf. J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cognitive architecture: A critical analysis*.

[33] Cf. S. PINKER, *The language instinct*.

[34] The entertainment industry, for example, relies heavily on this ability, e.g. in creating fantasy worlds to which we can nevertheless relate, e.g. because they express familiar "who does what to whom" relations.

[35] Cf. N. CHOMSKY, *Syntactic structures*.

[36] For a discussion on other aspects cf., e.g., F. VAN DER VELDE, *The neural blackboard theory of neuro-symbolic processing.*

[37] J.A. FODOR, Z.W. PYLYSHYN, *Connectionism and cog-*

nitive architecture: A critical analysis, p. 65 – italics by the authors.
[38] Cf. H. Rogers, *Theory of recursive functions and effective computability*.
[39] Cf. A.M. Turing, *On computable numbers, with an application to the Entscheidungsproblem*.
[40] Cf. A. Newell, *Unified theories of cognition*.
[41] Cf. H. Rogers, *Theory of recursive functions and effective computability*.
[42] Cf. W. James, *The principles of psychology*.
[43] Cf. J.A. Fodor, Z.W. Pylyshyn, *Connectionism and cognitive architecture: A critical analysis*.
[44] Cf. E.A. Jackson, *Perspectives of nonlinear dynamics*. As outlined further in F. van der Velde, *Computation and dissipative dynamical systems in neural networks for classification*.
[45] Cf., e.g., *IBM's Watson Supercomputer Destroys Humans in Jeopardy*, available at Endgadget-URL: https://www.youtube.com/watch?v=WFR3lOm_xhE (last visit: 22 March 2022).
[46] Cf. J.A. Fodor, Z.W. Pylyshyn, *Connectionism and cognitive architecture: A critical analysis.*
[47] W. James, *The principles of psychology*, p. 108 - italics by the author.
[48] Cf. E.R. Kandel, J.H. Schwartz, T.M. Jessell, S.A. Siegelbaum, A.J. Hudspeth (eds.), *Principles of neural science*.
[49] Cf. A. Amsel, M.E. Rashotte (eds.), *Mechanisms of adaptive behavior*.
[50] M. Shanahan, *Embodiment and the inner life*, p. 3.
[51] Cf. J.A. Fodor, Z.W. Pylyshyn, *Connectionism and cognitive architecture: A critical analysis*.
[52] Cf. F. van der Velde, *Computation and dissipative dynamical systems in neural networks for classification*.
[53] Cf. T.B. Brown, B. Mann, N. Ryder, M. Subbiah, I. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D.M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, *Language models are few-shot learners*.
[54] Cf. T. Kriete, D.C. Noelle, J.D. Cohen, R.C. O'Reilly, *Indirection and symbol-like processing in the prefrontal cortex*.
[55] Cf. D.J. Watts, S.H. Strogatz, *Collective dynamics of "small-world" networks*; M. Shanahan, *Embodiment and the inner life*.
[56] Cf. F. van der Velde, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*.
[57] Cf., e.g., M.D. Hauser, N. Chomsky, W.T. Fitch, *The faculty of language: What is it, who has it, and how did it evolve?*.
[58] Cf. J.A. Fodor, Z.W. Pylyshyn, *Connectionism and cognitive architecture.*
[59] Cf. A. Newell, *Unified theories of cognition*.
[60] Cf. F. van der Velde, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing.*
[61] Cf. A. Newell, *Unified theories of cognition*.
[62] Cf. P. Bloom, *How children learn the meaning of words*.
[63] Cf., e.g., A.G. Huth, W.A. de Heer, T.L. Griffiths, F.E. Theunissen, J.L. Gallant, *Natural speech reveals the semantic maps that tile human cerebral cortex*.
[64] F. van der Velde, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*.
[65] Cf. A. Newell, *Unified theories of cognition*.
[66] Cf. T. Kriete, D.C. Noelle, J.D. Cohen, R.C. O'Reilly, *Indirection and symbol-like processing in the prefrontal cortex*.
[67] Cf., e.g., A. Newell, *Unified theories of cognition*.
[68] Cf. D.O. Hebb, *The organisation of behavior: A neuropsychological theory*.
[69] Cf., e.g., A.G. Huth, W.A. de Heer, T.L. Griffiths, F.E. Theunissen, J.L. Gallant, *Natural speech reveals the semantic maps that tile human cerebral cortex*; M.A. Lambon-Ralph, E. Jefferies, K. Patterson, T.T. Rogers, *The neural and computational bases of semantic cognition*.
[70] Cf. F. van der Velde, J. Forth, D.S. Nazareth, G.A. Wiggins, *Linking neural and symbolic representation and processing of conceptual structures*.
[71] Cf., e.g., R.C. O'Reilly, J.W. Rudy, *Conjunctive representations in learning and memory: Principles of cortical and hippocampal function*.
[72] Cf. T. Kriete, D.C. Noelle, J.D. Cohen, R.C. O'Reilly, *Indirection and symbol-like processing in the prefrontal cortex*.
[73] Cf. D.J. Watts, S.H. Strogatz, *Collective dynamics of "small-world" networks*; M. Shanahan, *Embodiment and the inner life*.
[74] More details on the underlying mechanisms and examples can be found in, e.g., F. van der Velde, M. de Kamps, *Neural blackboard architectures of combinatorial structures in cognition*; F. van der Velde, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*; F. van der Velde, *The neural blackboard theory of neuro-symbolic processing: Logistics of access, connection paths and intrinsic structures*.
[75] Cf. P. Bloom, *How children learn the meaning of words*.
[76] Cf. F. van der Velde, M. de Kamps, *Neural blackboard architectures of combinatorial structures in cognition*.
[77] Cf. F. van der Velde, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*; F. van der Velde, *The neural blackboard theory of neuro-symbolic processing: Logistics of access, connection paths and intrinsic structures*.
[78] Cf. D.O. Hebb, *The organisation of behavior*.
[79] Cf. A.G. Huth, W.A. de Heer, T.L. Griffiths, F.E. Theunissen, J.L. Gallant, *Natural speech reveals the semantic maps that tile human cerebral cortex*.
[80] Cf. F. van der Velde, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*.
[81] Cf. S. Pinker, *The language instinct*. It is beyond the scope of this article to outline this further, for an overview cf., e.g., F. van der Velde, *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*.
[82] Cf. A. Newell, *Unified theories of cognition*.
[83] Cf. J.A. Fodor, Z.W. Pylyshyn, *Connectionism and cognitive architecture: A critical analysis*.

Literature

Amsel, A., Rashotte, M.E. (eds.) (1984). *Mechanisms of adaptive behavior: Clark L. Hull's theoretical pa-*

pers, with commentary, Columbia University Press, New York.

BLOOM, P. (2000). *How children learn the meaning of words*, MIT Press, Cambridge (MA).

BROWN, T.B., MANN, B., RYDER, N., SUBBIAH, M., KAPLAN, I., DHARIWAL, P., NEELAKANTAN, A., SHYAM, P., SASTRY, G., ASKELL, A., AGARWAL, S., HERBERT-VOSS, A., KRUEGER, G., HENIGHAN, T., CHILD, R., RAMESH, A., ZIEGLER, D.M., WU, J., WINTER, C., HESSE, C., CHEN, M., SIGLER, E., LITWIN, M., GRAY, S., CHESS, B., CLARK, J., BERNER, C., MCCANDLISH, S., RADFORD, A., SUTSKEVER, I., AMODEI, D. (2020). *Language models are few-shot learners*. In: «ArXiv», arXiv:2005.14165v4 - doi: 10.48550/arXiv.2005.14165, last revision: 22 July 2020.

BROWNING, J., LE CUN, Y. (2022). *What AI can tell us about intelligence*. In: «Noema», available at URL: https://www.noemamag.com/what-ai-can-tell-us-about-intelligence/

CHOMSKY, N. (1957). *Syntactic structures*, Mouton, The Hague.

FODOR, J.A., PYLYSHYN, Z.W. (1988). *Connectionism and cognitive architecture: A critical analysis*. In: S. PINKER, J. MEHLER (eds.), *Connections and symbols*, MIT Press, Cambridge (MA), pp. 73-193.

HANSON, N.R. (1958). *Patterns of discovery: An inquiry into the conceptual foundations of science*, Cambridge University Press, Cambridge.

HAUSER, M.D., CHOMSKY, N., FITCH, W.T. (2002). *The faculty of language: What is it, who has it, and how did it evolve?*. In: «Science», vol. CCXCVIII, n. 5598, pp. 1569-1579.

HEBB, D.O. (1949). *The organisation of behavior: A neuropsychological theory*, Wiley, New York.

HOTHERSALL, D. (2004). *History of psychology*, McGraw-Hill, Boston, 4th edition.

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*. In: «Nature», vol. DXXXII, n. 7600, pp. 453-458.

JACKSON, E.A. (1991). *Perspectives of nonlinear dynamics*, 2 voll., Cambridge University Press, Cambridge.

JAMES, W. (1950). *The principles of psychology* (1890), 2 voll., Dover Publications.

KANDEL, E.R., SCHWARTZ, J.H., JESSELL, T.M., SIEGELBAUM, S.A., HUDSPETH, A.J. (eds.) (2013). *Principles of neural science*, McGraw-Hill, New York, 5th edition.

KÖHLER, W. (1965). *On the insight of apes* (1917). In: R.J. HERRNSTEIN, E.G. BORING (eds.), *A source book in the history of psychology*, Harvard University Press, Cambridge (MA), pp. 569-578.

KRIETE, T., NOELLE, D.C., COHEN, J.D., O'REILLY, R.C. (2013). *Indirection and symbol-like processing in the prefrontal cortex*. In: «Proceedings of the Academy of Sciences of the United States of America», vol. CX, n. 41, pp. 16390-16395.

KUHN, T.S. (1970). *The structure of scientific revolutions* (1962), University of Chicago Press, Chicago, 2nd edition.

RALPH, M.A.L., JEFFERIES, E., PATTERSON, K., ROGERS, T.T. (2017). *The neural and computational bases of semantic cognition*. In: «Nature Reviews Neuroscience», vol. XVIII, n. 1, pp. 42–55.

MARCUS, G. (2020). *The next decade in AI: Four steps towards robust Artificial Intelligence*. In: «ArXiv», arXiv:2002.06177 – doi: 10.48550/arXiv.2002.06177 - last revision 2020, February 19th.

MARCUS, G., DAVIS, E. (2020). *GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about*. In: «MIT Technological Review», published: 22 August 2020, available at URL: https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/.

NEWELL, A. (1990). *Unified theories of cognition*, Harvard University Press, Cambridge (MA).

O'REILLY, R.C., RUDY, J.W. (2001). *Conjunctive representations in learning and memory: Principles of cortical and hippocampal function*. In: «Psychological Review», vol. CVIII, n. 2, pp. 311-345.

PAVLOV, I.P. (1965) *On conditioned reflexes* (1904). In: R.J. HERRNSTEIN, E.G. BORING (eds.), *A source book in the history of psychology*, Harvard University Press, Cambridge (MA), pp. 564-569.

PINKER, S. (1994). *The language instinct*, Penguin, London.

PYLYSHYN, Z.W. (1984). *Computation and cognition: Toward a foundation for cognitive science*, MIT Press, Cambridge (MA).

ROGERS, H. (1988). *Theory of recursive functions and effective computability*, MIT Press, Cambridge (MA).

ROSENBLATT, F. (1958). *The perceptron: A probabilistic model of information storage and organization in the brain*. In: «Psychological Review», vol. LXV, n. 6, pp. 386-408.

SHANAHAN, M. (2010). *Embodiment and the inner life*, Oxford University Press, Oxford.

SILVER, D., SCHRITTWIESER, J., SIMONYAN, K., ANTONOGLOU, I., HUANG, A., GUEZ, A., HUBERT, T., BAKER, L., LAI, M., BOLTON, A., CHEN, Y., LILLICRAP, T., HUI, F., SIFRE, L., VAN DER DRIESSCHE, G., GRAEPEL, T, HASSABIS, D. (2017). *Mastering the game of Go without human knowledge*. In: «Nature», vol. DL, n. 7676, pp. 354-359.

SKINNER, B.F. (1957). *Verbal behavior*, Appleton-Century-Crofts, New York.

THORNE, B.M., HENLEY, T.B. (2001). *Connections in the history and systems of psychology*, Houghton Mifflin, Boston.

TURING, A.M. (1937). *On computable numbers, with an application to the Entscheidungsproblem*. In: «Proceedings of the London Mathematical Society», vol. XLII, S2, n. 1, pp. 230-265.

VAN DER VELDE, F. (in press). *The neural blackboard theory of neuro-symbolic processing: Logistics of access, connection paths and intrinsic structures*. In: P. HITZLER, K. SARKER, A. EBERHART (eds.), *Compendium of neuro-symbolic artificial intelligence*, IOS Press, Amsterdam.

VAN DER VELDE, F. (2022). *Towards a neural architecture of language: Deep learning versus logistics of access in neural architectures for compositional processing*. In: «ArXiv», arxiv.org/abs/2210.10543 – doi: 10.48550/arXiv.2210.10543.

VAN DER VELDE, F. (2015). *Computation and dissipative dynamical systems in neural networks for classification*. In: «Pattern Recognition Letters», vol. LXIV, pp. 44-52.

VAN DER VELDE, F., DE KAMPS, M. (2006). *Neural blackboard architectures of combinatorial structures in cognition*. In: «Behavioral and Brain Sciences»,

vol. XXIX, n. 1, pp. 37-70.

VAN DER VELDE, F., FORTH, J., NAZARETH, D.S., WIGGINS, G.A. (2017). *Linking neural and symbolic representation and processing of conceptual structures*. In: «Frontiers in Psychology», vol. VIII, Art. Nr. 1297 - doi: 10.3389/fpsyg.2017.01297.

WATSON, J.B. (1913). *Psychology as the behaviorist views it*. In: «Psychological Review», vol. XX, n. 2, pp. 158-177.

WATSON, J.B. (1924). *Behaviorism*, The People's Institute Publishing Co., New York.

WATTS, D.J., STROGATZ, S.H. (1998). *Collective dynamics of "small-world" networks*. In: «Nature», vol. CCCXCIII, n. 6684, pp. 440-442.