Language as a tool for simplified cognition in the semantic Bayesian brain

THE COMPUTATIONAL MIND AND ITS CRITICS
LECTURER: JULIAN KIVERSTEIN

Logic and Language Group
Department of Philosophy
University of Amsterdam

Mick de Neeve
mick@live.nl
REG. 9305017

February 2, 2015

Abstract Shagrir's semantic-computational perspective restricts the notion of computation to brains and familiar computing devices, but only demonstrably applies to simple actions and not to higher cognition. This is harder to capture in terms of precise functionality and cannot generally be claimed to inherit a logical semantics, leading to a gap in the semantic perspective. But another, linguistic sense of 'semantics' can be applied to Dreyfus's post-logicist higher cognition view of expertise, by considering it from Clark's perspective of the brain as a Bayesian prediction machine, augmented with previous work by Clark on language as an artefact. That imports semantics into higher, not strictly linguistic or logicist cognition, and results in an account of language as a learning foundation that persists to avoid overfitting.

1 Introduction

Cognitive science is founded on the idea that the human mind can be understood as a computer. During the second half of the twentieth century, this analogy was pretty much unquestioned. Within the so-called logicist tradition, the mind was taken as an ideal not unlike frictionless motion in physics. The approach is characterized by rigid rational central control, with problematic cases that did not quite fit such as commonsense reasoning (e.g. the infamous frame problem) being deemed peculiarities to be dealt with separately, later, or not at all.

It had been known for a while that the brain is filled with vast amounts of 'dumb' nerve cells, for which McCullogh and Pitts had come up with mathematical models as early as 1943 [MP43]. However, computers can similarly be considered dumb at the logic gate level of bits, and so the idea was that a computational process which the mind implements by firing neurons could be implemented on a computer which manipulates ones and zeros. From the logicist perspective, the mind thus runs programs and is trivially a computer. But logicism has not persevered, and various connectionist and probabilistic accounts of cognition have come to replace it. At the same time, a broad computationalist perspective remains, which has consequently become less straightforward to defend, most notably because of the so-called pancomputationalism problem.

Pancomputationalism is the problem of defining computation too widely. For instance, a definition in terms of input and output states with some sort of transformation between them, forces one to consider practically anything as a computational system, including stomachs and planetary systems. As Piccinini ([Pic12], section 3.1) points out, this is unfair to the field of computer science with its extensive efforts in designing reliably operating computational systems. Moreover, if anything can be a computer, the computational perspective in cognitive science becomes meaningless. The solution proposed by Shagrir is to consider a computational theory of the capacities of the brain as one where the objective is to figure out how a 'semantic' task is performed ([Sha06], p.394).

According to Fodor, however, mental processes can only be considered computational if they are formal, meaning that if anything is represented in the mind, the mind does not have access to what it is about but only has the form of that representation to go on ([Fod80], p.65). Consequently anything called 'semantic' with regard to a representation would have to be mirrored by the formal (i.e. the internal) properties of the representation. This view is echoed by Chalmers, whose slogan in this respect is: *"If you take care of the syntax, the semantics takes care of itself."* Besides, Chalmers considers the notion of semantic content not sufficiently well-understood to play any robust sort of role in the foundations of cognitive science ([Cha94] p.9).

But can semantics really be considered as 'taking care of itself', given that it has to come from somewhere? What mental states are about is not somehow locked up in the brain, but originates in perception, action, and communication. So a perfect, closed harmony between operation, form or syntax, and meaning, content or semantics cannot be assumed, but this should not prevent semantics from playing a fundamental role in cognitive science.

However, if the notion of semantics is widened to include Shagrir's neural representations, then in order to have a clear 'aboutness', neural patterns could seemingly not concern much more than the simple situations and actions he describes. For as the account of expertise by Dreyfus suggests, aboutness of higher cognition can be difficult if not impossible to pinpoint. The solution proposed here is to associate this sort of higher cognition with language, even if it is not always expressly linguistic.

1.1 Overview

To this end, the following section will first of all address Clark's views on language as a tool to augment cerebral computation, followed by Shagrir's semantics as discriminator for computing processes at the basic level of perception and action. Next up is Dreyfus's criticism of expert systems, which is brought into accordance with Clark's views on the predictive Bayesian brain. Finally, Clark's account of language is related to the aim of biasing prediction models towards simplicity.

It is ultimately argued that this bias is necessary to prevent the Dreyfus type of expert from overfitting the data encountered in the world, and that it is language which facilitates this. It is thus via this foundational role of language that higher cognition can be understood as semantic, bypassing the need for having to understand it precisely along the lines of Shagrir.

2 Language, semantics, and computation

2.1 Private language and serial computation

In *Magic Words: How Language Augments Human Computation* [Cla98], Clark puts forward a view of language as a tool which augments human computation, and notably as a means to assist the brain in the planning and execution of tasks. Although language developed as a vehicle for communication, Clark points to studies from psychology (cf. Vygotsky [Vyg34]), in which the private use of language was shown to play an important role in the coordination of complex tasks by subjects, who would state sequences of goals and subgoals to themselves and follow them. They were moreover found to rehearse these goals in sentences, and scrutinize and alter these if needed.

This highlights two points. Firstly, in terms of semantics, Clark points out that words need to constitute a procedural 'code' which minimizes sensitivity to context (ibid, p.12), so that they may be scrutinized from various angles (in the mind) without their meaning being altered. Secondly, this planning of procedures with goals and subgoals constitutes a mechanisation of the process of thought which is arguably at the basis of the development of computers and programs, i.e. the very lens through which the mind is viewed in the computational paradigm of cognitive science.[1] In other words, while the brain operates in a (massively) parallel fashion, the standard Von Neumann architecture computer is a serial, or sequential machine, which has the equally sequential device of stable-semantics human language at its basis.

So language can in this sense be viewed as a simplificational tool that acts as a container to sequentially capture and summarize what the brain has come up with in its parallel operations. And when such language-encapsulated actions of the mind are (logicistically) modeled on a computer, the computer can be said to be doing the sort of thing that gave rise to its very invention. So conversely when the brain does them, it is quite uncontroversially computation, and semantic, with semantics used in the classic linguistics-inherited sense.

[1] This causes something of a loop with respect to questions of scientific foundation, which has inspired Coolen [Coo92] to argue that the computer must first and foremost be understood from the perspective of man, rather than the other way round.

2.2 Shagrir's semantic computation

Shagrir defines a semantic task as one that is specified in terms of representational content ([Sha06], p.403). He cites a neural networks-based study by Shadmehr and Wise [SW05] about the computation of motor commands for the hand in order to grab an object in the visual field. This, the authors point out, involves transforming representations of hand and object locations to one that represents the difference between them.

Within the Shadmehr and Wise study, Shagrir focuses on the question how the brain works out the difference between object and 'end-effector' - e.g. the hand. The problem is described in terms of computing a difference vector given the activation vectors that represent the end-effector and the object, respectively.

According to Shadmehr and Wise, the brain computes the position of the hand from information about the angles of the joints, and the location of the object from information about the orientation of the eye and the neural mapping of the retinal image of the target. The description makes it clear that the authors are concerned with representations that are quite literally about something, in virtue of which Shagrir terms the computation 'semantic'.

But that does not as such say anything about the semanticity of higher cognition. So what to make of cases which the traditional computationalist perspective cannot quite cover?

2.3 Dreyfus's beginners and experts

In *From Socrates to Expert Systems: The Limits of Calculative Rationality* [Dre87], Dreyfus critises the use of reasoning models in the field of expert systems. He argues that experts 'do what works' rather than 'solve problems'.[2] Dreyfus does not view 'the expert' as a static concept, but distinguishes between various stages on the path towards expertise. Broadly speaking, a beginner starts out by reasoning with explicit rules about generalities, but increasingly, problem solving recedes in favour of the ability to recognize and act on specific cases. This runs counter to the view of learning as generalization, and presents a problem for knowledge engineers who try to extract rules from experts. But the lack of explicit rules also obscures what the expert's knowledge viz. cognition is actually about, and hence complicates a semantic-computational account.

Awareness of the rule extraction problem goes back as far as Socrates, notes Dreyfus, who also points out that it was even known to knowledge engineering pioneers. He quotes Feigenbaum [FM83], who described situations where experts claimed to simply rely on having seen sufficient amounts of cases. *"At this point, knowledge threatens to become ten thousand special cases"*, Feigenbaum complained.

But Dreyfus does not so much complain as argue that this is simply the way

[2] In a later version [Dre02] it is actually denied that expert skills are representational, but Dreyfus is talking about representations in the logicist sense which is not the same as Shagrir's neural perspective.

learning and expertise work. Consequently, he states, to push the expert to explicating rules is to push him back down to the beginner level ([Dre87], p.30).

As Churchland points out, 'processing' and 'memory' are not separate in the brain like in a computer: *"In the biological brain, to engage in any computational transformation simply is to deploy whatever knowledge it has accumulated"* ([Chu05], p7). But if processing and memory are so intertwined, then apparently somewhere, the distinction between syntax and semantics vanishes, and expertise becomes non-linguistic. Presumably this is the point at which logicism fails, and beyond which the expert, rather than 'solve problems', just 'does what works'.

So does language completely disappear from experts? In the next section, it is argued that gaining expertise can be understood as data compression, using Clark's ideas on the Bayesian brain. Knowledge may become increasingly non-linguistic, but is still based on language, and the classifying, dimensionality-reducing meanings of words endorsed earlier by Clark then helps to bias compression towards simpler models.

3 The Bayesian brain, cognition, and language

"The whole function of the brain is summed up in: error correction." This 1960s quote of the British psychiatrist Ashby is the opening line of Clark's *Whatever next? Predictive brains, situated agents, and the future of cognitive science* [Cla13]. It sums up the basic view of the paper: to posit not logical, but Bayesian probabilistic inference as the central operation of the mind. This can be associated with what is known about the operations of neurons, but also appears relevant to higher reasoning. This is a significant step forward compared to the logistic paradigm, which was faced with a gap between high-level rationality and low-level dumbness.

In the so-called predictive coding view of the brain, perception is cast in terms of data compression. That is, the brain does not grasp all that is going on, but encodes a short description - a hypothesis - and processes mismatches of this when predicting. As a not-very-cerebral compression example, consider a string consisting of a thousand zeros and a one. A short description is *'a thousand zeros'*, and the mismatch, or error, is *'and a one'* (which is more or less how computers would compress such a string, taking much less space than to completely write it out). For the brain, the associated prediction would be that a thousand zeros are to be observed. It is then the main business of the brain to propagate errors so as to improve future predictions. In case that missing *'and a one'* impairs actions, it is the only thing of concern.

More cerebrally, consider the Shadmehr and Wise task from this perspective. In their study and in Shagrir's presentation of it, grabbing the object is considered a done deal: the brain is in a trained state for it. But viewed as predictive coding, the underlying task is predicting input signals from the visual system, and also the motor commands hypothesized to minimize any mismatch between the predictions and the actual inputs, i.e. the commands that fulfil the predictions.

The brain thus predicts the perception of the world as well as of the body's configuration, in other words, proprioception. As Clark points out, the error in predicting the latter could be considered as actually constituting the motor commands (ibid, p.186). On that view, the brain would be applying its trained 'static' state in the Shadmehr and Wise mode as the broad hypothesis (the short description) for action (in this case, grabbing the object), while continuously applying corrections where things do not fit the learned picture (the error).

The process of acting at one point in time then becomes the process of learning in order to act more effectively later on, meaning that the error must not only generate (corrections of) motor commands, but become incorporated in the broad hypothesis. The goal of learning being the capacity to maximize posterior probabilities of generating effective actions - i.e. to maximize expected utility of actions - errors should come to be encoded as updated prior probabilities about the world. But maximization of expected utility through error correction is not the end of the story; it is only a goal in case some action is called for. Arguably, data compression, or the simplification of the world, is the wider goal, for which the brain needs ways to establish a clear bias towards simple models of the world.

If maximization of expected utility at any moment were the whole story, there could in principle be two ways to go when picking a model of the world. One would be towards overfitting, and the other towards underfitting the data. But as pointed out in Grünwald ([Grü98], ch.1, section 1.5), both have tended to be viewed as equally undesirable.[3] However, there is an argument for favouring underfitting. When the types of models to choose from, or model class, is very large,picking the model with the smallest error will tend to overfit the observed data, while in case of a very small model class, the one with the smallest error will tend to underfit. In the first case, the average error on the observed data will be a very poor indicator for the error likely to result on new data, while in the second case, this average error will predict future errors well. But then, given that the brain needs to predict motor commands, which as indicated may in fact constitute prediction errors themselves, the brain will have to be good at predicting errors. Hence to do this well there must be a bias towards underfitting.

Below, it is argued that language is a tool for achieving this, and that consequently Drefus's expert uses it even if not all actions can always be explained in terms of rules expressed in language.

3.1 Language, meaning and expertise

In section 2.1, Clark's views on private language use for task planning are highlighted. He stresses the need for stable word meanings, or minimized context sensitivity. But of course this is the same language as the public one, and thus equally suited for instructing beginners how to perform those tasks. From this perspective, the beginner's brain computes semantically by using language to select rules and plan actions, and learns by generalizing. In terms of model

[3]That is, in the literature on Bayesian networks and probabilistic reasoning.

selection, and taking Clark's views on the dimensionality reducing function of words into account, the beginner's model will arguably strongly underfit the data observed so far.

But as the Dreyfus type of beginner progresses, distinctions will have to be dealt with which cannot be properly captured by words. An example he gives for the stage of 'Advanced Beginner' ([Dre02], §II.2) is a driver learning to recognize engine sound cues for shifting gear. Increasingly, skills move beyond the realm of linguistic descriptivity into that of situational affordances. At the expert level, given a set of such situations plus actions, one might try to capture these linguistically, but the result should then be expected to be highly specific, or, in model selection terms, strongly overfitting and thus quite useless for handling novel situations.

This might seem puzzling, since has language not just been positioned as a generalizational device? Indeed, but by considering it as part of the 'more cognitive realm' (Clark [Cla13], p.190), language is about capturing regularities at larger temporal and spatial scales. Its virtues with regard to context-insensitivity are prone to being lost when applied to the wrong scale. Semantically speaking, this means term-denotational sets could be shrunk to the point of singletons. Resulting descriptions in terms of rules might then be expected to become too dependent on context, and consequently lead to overfitting models.

However, this may only be the case when language is used in its public instructional role. In the 'Competence' stage of progress from beginner to expert ([Dre02], §II.3), Dreyfus argues that the learner will now devise plans, and seek new rules and reasoning procedures. Naturally, this will occur in a private language use context. But then, note one of Clark's uses of language as a tool to assist the computing brain ([Cla98], p.7) is environmental simplification.

On this view, language helps to simplify the surroundings humans have to negotiate, by labeling things out there that have to be dealt with. But Clark particularly emphasizes that this leads to a simplification of learning environments. He describes the advantage of language as a strategy that provides the environment with a stability which reduces computational demands on the brain. So the private use of language applied by the learner when planning and reasoning at Dreyfus's stage of 'competence' should reduce the size of the model class from which a current model of the environment is to be chosen. Consequently, a bias towards underfitting should result, and plausibly, this persists into the stage of expertise.

The expert's brain can now be said to compute semantically in two ways. Firstly, at the lower level of action, it does so in the Shadmehr and Wise sense. Secondly, at the higher level of cognition, even when in the non-linguistic expertise mode, it does so based on public linguistic instruction as well as private linguistic planning and reasoning, both of which utilize generalizing and simplifying features of natural language semantics.

4 Discussion

The previous sections indicates how language can be considered as a private planning and reasoning device - which arguably provided some of the inspirations behind classic computing machinery - but also as a tool to steer neural computation towards effectivity. This highlights one way in which the 'Bayesian brain' perspective advocated by Clark can move towards a unifying view of perception, action and cognition ([Cla13], section 5.2). Human learning environments have become, as Clark puts it, 'thoroughly artificial', which makes such a unifying theory harder to achieve, but language and the structuring capacities it provides arguably constitute an important means by which these learning contexts have become the way they are.

But another way to include the influence of language on 'higher cognition' is to consider attentive thought from the prediction perspective (which would incidentally aid the unification effort more directly). On such an account, the business of the thinking brain is to predict its own thoughts. But then, notwithstanding the infinity of linguistic combinations, the generalizational features of language may in fact help to reduce the number of possible thoughts, making them easier to predict. Although this view would assume some form of a 'language of thought', it may end up helping to explain those cases where one is lost for words, or words do not suffice to express thoughts, since thoughts would be subject to compression and may end up as hard to access as 'expert rules'.

5 Concluding remarks

In this essay, the semantic view of the computational mind has been extended from aboutness of lower level neural networks to linguistic meaning, relating to higher level cognition as viewed from a Bayesian prediction perspective. To this end, it has been argued that language plays a crucial role in catering for the brain's ability to select models of the world that are sufficiently complex to negotiate complex environments, without being too specific for novel situations.

References

[Cha94] D. Chalmers. A Computational Foundation for the Study of Cognition. *Minds and Machines*, 4, 1994.

[Chu05] P. Churchland. Functionalism at forty: A critical retrospective. *Journal of Philosophy*, 102(1):33–50, 2005.

[Cla98] A. Clark. Magic Words: How Language Augments Human Computation. In P. Carruthers and J. Boucher, editors, *Language and Thought: Interdisciplinary Themes*, pages 162–183. Cambridge University Press, 1998.

[Cla13] A. Clark. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36, 2013.

[Coo92] M. Coolen. *De machine voorbij; over het zelfbegrip van de mens in het tijdperk van de informatietechniek*. PhD thesis, University of Amsterdam, 1992. English title: The machine and beyond; on the self-concept of man in the age of information technology.

[Dre87] H. Dreyfus. From Socrates to Expert Systems: The Limits of Calculative Rationality. *Bulletin of the American Academy of Arts and Sciences*, 40(4):15–31, 1987.

[Dre02] H. Dreyfus. Intelligence without representation. *Phenomenology and the Cognitive Sciences*, 1(4):367–383, 2002.

[FM83] E. Feigenbaum and P. McCorduck. *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World*. Addison Wesley, 1983.

[Fod80] J. Fodor. Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*, 3(1):63–73, 1980.

[Grü98] P. Grünwald. *The Minimum Description Length Principle and Reasoning under Uncertainty*. PhD thesis, University of Amsterdam, 1998.

[MP43] W. McCulloch and W. Pitts. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

[Pic12] G. Piccinini. Computationalism. In *Oxford Handbook of Philosophy and Cognitive Science*, chapter 9. Oxford University Press, 2012.

[Sha06] O. Shagrir. Why we view the brain as a computer. *Synthese*, 153:393–416, 2006.

[SW05] R. Shadmehr and S. Wise. *The computational neurobiology of reaching and pointing: A foundation for motor learning*. MIT Press, 2005.

[Vyg34] L. Vygotsky. *Thought and Language*. State Social-Economic Press, Moscow, 1934. Translation from Russian by E. Hanfmann and G. Vakar, MIT Press, 1962.