Can memes explain the birth of comprehension?

Paweł Grabarczyk
IT University of Copenhagen,
University of Lodz
pagrab @ gmail.com

Received 16 April 2018; accepted 5 July 2019; published 31 December 2019

Abstract

The paper is a polemic with Daniel Dennett’s account of memes presented in his latest book *From Bacteria to Bach and Back: The Evolution of Minds*. This latest attempt at explaining the idea of memes ties it to the recent advances in Artificial Intelligence and tries to show how the mechanism of meme transmission affected the birth of human intelligence (which Dennett calls “comprehension”). I argue that even though the new version of the theory is not without its merits, it comes with two fatal flaws. First of all, even though on the surface Dennett talks only about words, gestures signs etc., he seems to smuggle in an additional notion of linguistic meaning which remains unexplained. Second of all, he does not provide identity conditions for memes which makes the whole idea of their transmission completely vague.

Keywords: memes; comprehension; artificial intelligence; meaning; Dennett

1. From competence to comprehension.

Daniel Dennett’s *From Bacteria to Bach and Back* consists of two importantly different parts. The first part, one which I have no issues with, explains the key difference between comprehension and competence. Even if giving a precise definition of both of these concepts may end up being quite difficult, the distinction itself is rather intuitive, because it mirrors the “user experience” within our brains. When I make use of some of my competences - for example, when I look for a word that I need, my experience reminds me of using a computer. I feel that I need a better word and then, sooner or later, the word pops up in my mind. I find that my mind is literally “jumping to the conclusion”. I have no idea as to what happened between these two moments, even though I consider myself a competent language user. The difference

---

1 Update: 2 February 2020.
between these moments and the “a-ha” feeling which accompanies comprehension is hard to ignore. Whenever I truly comprehend something, it is the opposite of “jumping to the conclusion”. Not only do I feel that I know the answer to a given question, but I also know how I arrived at it. This difference may, for all that we know, turn out to be a phenomenological illusion, but it definitely makes Dennett’s distinction intuitive. In the remaining part of the current section, I explore the distinction between competence and comprehension further. In Section 2 I proceed to the most problematic aspect of the book – the theory of memes and their role in Dennett’s argumentation. As I claim in Section 3, the description of memes found in From Bacteria to Bach and Back leaves much to be desired as it is difficult to put memes in the context of other existing philosophical posits, such as meanings. Section 4 looks at the role of memes in Dennett’s solution to the problem of the emergence of comprehension and suggests a more coherent picture.

It could be argued that the difference between competence and comprehension is now timelier than ever. The rise of new forms of AI and machine learning resulted in many impressively competent machines that seem to have no comprehension whatsoever. This shift in the AI paradigm is the reason for the title of the book as current artificial intelligence is more reminiscent of hive minds and similar creations of nature than the super-intelligence predicted by many science-fiction novels. There is a great sense of irony here, which Dennett exploits, as it may seem that our form of intelligence, which is, at least according to our own image of it, based on comprehension, appears more like a small bump (or maybe even a detour) on the evolutionary path and not the crowning achievement of evolution we thought it to be. Nature produced more and more competent systems which resulted in the emergence of comprehension. Systems equipped with comprehension accelerated the evolution of intelligence through cultural evolution and technology but the end result of all of this effort is that sheer competence took over again. This picture makes comprehension even more mysterious than it was and raises the question of its actual utility.

This change in paradigm can be best seen in a microscale once we follow the development of Alpha Zero - an AI computer program created by the company Deep Mind to show the possibility of learning and mastering any existing game. Alpha Zero's role in our understanding of classic games, such as Chess, Shogi or Go can be quite instructive. Deep Mind's AI differs from the classic approaches to artificial players in that it gains its competence completely autonomously. Classic game playing systems functioned as the embodiment of our comprehension of games as their algorithms reflected our understanding of games' rules and the best strategies discovered by humans throughout history. However, Alpha Zero starts without any assumptions about the games it tries to learn and trains itself by observation and by playing against other instances of itself.

The analysis of Alpha Zero's matches already transformed the way humans play classic games, such as Go (Lee, 2017). Professional players obtained a lot of useful know-how about the game, even though we do not get the rationale behind Alpha Zero's choices. Importantly, this is very different from the classic science-fiction visions of AI in which the lack of explanation of machine decisions was typically tied to the fact that the super-intelligence surpassed our

---

2 Dennett discusses the earlier version of the program - AlphaGo in (Dennett, 2017, p. 418).
Can memes explain the birth of comprehension?

reasoning to the extent where we would have simply been unable to understand its rationale. The case of Alpha Zero is much more prosaic. There is simply no rationale behind its competence. At least not in the sense we understood rationale up to this point. Alpha Zero is more similar to a witch-doctor than to a doctor. It can show us what we should do to succeed, but it cannot explain why this strategy is successful. Even if, after additional analysis, we can understand the reasons why the strategy works, it has to be understood independently as the only lesson Alpha Zero could give us is that the strategy increased the probability of winning. The crucial point is that the process of understanding the rationale behind this strategy couldn't even be called "reverse engineering" in a standard sense. Our speculations did not recreate the mechanism behind Alpha Zero's decision but rather invented an explanation of a different type - one that gives us the much desired "a-ha" effect. Its emulation, not simulation.

The case of Alpha Zero is interesting because it may give us a glimpse of our future relations with competent artificial intelligence. Games – no matter whether simple (like Chess) or complex (like Go) – are rule-based systems that are isolated from their environment. Because of that, they can function as models or idealizations of more complex, open systems such as science or natural languages. The lesson we learned from Alpha Zero and, especially, our inability to understand its decision processes is profound and should probably curb our enthusiasm about the new paradigm of AI. That the expectations of some of the apologists of new AI need taming can be best seen in the case of ultra-optimism towards a new paradigm of science, best evidenced by the following quote from a popular science magazine:

There is now a better way. Petabytes allow us to say: ‘Correlation is enough.’...We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot . . . Correlation supersedes causation, and science can advance even without coherent models, unified theories, or any mechanism explanation at all. There’s no reason to cling to our old ways. (Anderson, 2008).

Dennett is well aware of the naive optimism associated with this “black box” approach to science and warns against uncritical thinking about the future filled with merely competent artificial agents. Even if there is no doubt that we will continue on this path and develop even more advanced competent AI, we should always be aware of the specific way it functions. Even though Dennett does not use this expression, we could say that we should refrain from employing a “comprehendingal stance” towards machines lacking comprehension.

The most important result of this “diagnostic” part of the book is that it makes a sharp distinction between comprehension and competence which leads to a key question: if merely competent agents that we create can match or even quite possibly surpass our cognitive abilities, then what exactly do we need comprehension for? What role does it play in our cognition and, more importantly, how did it originate in our evolution?

---

3 At least in the conceptual sense. Dennett does not suggest that comprehension popped out of competence in its fully developed form.
2. Memes to the rescue

To answer these questions, Dennett reintroduces the theory of memes which gets possibly the most thorough explanation since its introduction in (Dawkins, 1989). Dennett rightly points out that the theory was somewhat unlucky. In the beginning, it was not treated seriously as it was, at best, considered to be a far-fetched metaphor with low explanatory value. Some researchers even thought of it as dangerously close to pseudo-science (Benitez-Bribiesca, 2001). The current context in which the word "meme" resurged does not help its case, because it has been popularized in online communities, where it is typically used to denote a combination of pictures and short messages shared between the users of social networks as well as set phrases and jokes understood and quickly recognized by community members. On the surface, this modern usage does not have much of a connection with the original idea, but it seems to me that it is not completely unwarranted and can be simply treated as a bastardized form of Dawkins' memes.

I believe that the connection between Internet memes and memes as presented by Dennett is even stronger, because, ironically, the explanation of meme transfer provided in Bacteria to Bach and Back suits Internet memes better than other instances of what Dennett classifies as memes. The way memes are proliferated online is often non-reflexive and almost automatic. They are "shared" on social networks, they start to function in ordinary speech and their users may sometimes be hard-pressed to explain why exactly they pass them further. They simply seemed attractive at the moment. This mechanism is very well in line with what Dennett sees as competence and, as he is well aware, it can be very easily exploited by clever meme manipulators (Dennett, 2017, p. 392). We all know this mechanism from advertising and politics, but some of the recent usages of memes can show an even darker side of this unreflective replication. As pointed out in (Salter, 2018), (Hawley, 2017) and (Nagle, 2017), memes played an important role in the recent rise of the so-called "alt-right" movement which used online technologies to spread highly problematic ideology using the vehicle of seemingly non-harmful humor. It can be argued that the success of this technique depends on the fact that memes are prone to be transmitted without comprehending their true meaning and the context in which they have been used. The most prominent example of this dangerous mechanism can be found in so-called "dog whistles" – messages and symbols which are coded in such a way that their users transmit them without knowing their true meaning (López, 2015). If developed further, Dennett's understanding of memes seems to be a very promising tool for the analysis of these phenomena.

Unfortunately for Daniel Dennett, he wishes this theory to play a much bigger part as it is supposed to bridge the gap between mere competence and comprehension which he described in the first part of the book. This is something, which, in my opinion, the theory simply does not deliver. This specific problem with memetics will be the main subject of the rest of my paper. The criticism I present below should not be thus perceived as an overarching commentary on the theory, but rather on its single aspect – its potential to explain the way comprehension emerged from competence. Starting with the analogy with genes I will be looking at the way memes and their transmission are explained in the book. Then, I am going to analyze if this mechanism throws any light on the relation between competence and comprehension.
I finish with a suggestion as to how connecting Dennett's account with Robert Brandom’s inferentialism could shed some additional light on the matter.

I believe that, even though the analogy between genes and memes may initially seem attractive and perhaps illuminating, it eventually ends up being rather rough and somewhat misleading. Already at the time when genes were only theoretical posits, they functioned as postulated explanations of something observable, that is of the hereditary patterns. Genes played the role of unobserved properties responsible for the regularity of the transmission of observable properties. What it means is that they literally filled the gaps in our understanding of biological phenomena.

If we follow this analogy and look at memes, we could also start with some observable phenomena – words being repeated, ideas being shared, sayings being carried over through generations, tunes being popularized and so on. But here is where the analogy breaks down. If the analogy was to be strict, we would have to say that all of these instances of human behavior were copied because of the existence of memes – invisible entities responsible for the transmission (just as the invisible genes are responsible for the transmission of visible properties of organisms). But this is not the case in Dennett's account. Memes are not supposed to be the extra posits responsible for the transmission. They are not the unobserved fillers of our epistemic gaps. They are an overarching category that connects all of the instances of behavior being transmitted. In contrast to genes, memes function as an umbrella term for the properties or phenomena we already very well know. They function as a common label for words, meanings, ideas, customs, concepts or technologies. At best, they are just a handy name for any behavior that is transmitted or replicated for obscure reasons.

3. How does meaning relate to memes?

The problem is that most of what counts as memes – ideas and concepts being the most prominent example of this – is notoriously vague and hard to explain by means of philosophical theories. Meme transmission demands to be explained, rather than just grouped together under a new moniker. In other words, it’s not that memes rebrand things that other theories can explain. They seem to rebrand things other theories cannot explain.

This criticism may seem a little bit too harsh. Is it not the case that the sheer grouping of phenomena so diverse as words, dances, and technologies lets us learn something useful, namely that the mechanism of their transmission is homogenous? Even if this were true, however, the only thing memetics succeeds at is that it embellishes our lack of understanding of this mechanism.

This issue is most visible in the case of the transmission of words. Words are Dennett’s favorite case of memes, “the best example of memes” (Dennett, 2017, p. 176). The reason for this choice is quite simple. Philosophers may argue over the existence of meanings, concepts, ideas, etc., but they hardly argue over the existence of words.

4 Although I am not alone when it comes to this concern - see (Mayr, 1997).
The theory I am defending declares, nonmetaphorically and literally and without caveat, that words are memes that can be pronounced. Memes exist because words are memes, and words exist, and so do other ways of doing things that are transmitted nongenetically. (Dennett, 2017, p. 224).

The problem is that, despite these declarations, whenever Dennett talks about the transmission of words, he clearly refers to the transfer of the words with some meaning assigned to them which makes his theory much less clear than it was supposed to be. There are indeed no doubts as to whether words exist. It is also quite true that words are often repeated, copied and transferred between users across the borders and the millennia. The transmission process can often introduce small changes – words are pronounced differently, their spelling may change, etc. On the face of it, this plays into the analogy with genetic evolution. Dennett suggests that memes, just as genes, are sometimes copied with small alterations. This, of course, strengthens his analogy but it also makes it very clear that what is badly needed here is a differentiation between copying a word and copying the word with its meaning. The process of copying words from language to language has its quirks as words are very often copied with errors which lead to alterations. This is how differences between languages belonging to the same family often originate. But the picture gets complicated once we realize that words are often copied or borrowed without retaining their original meaning. This is how false friends are born.

It is also possible for people to intentionally borrow a word-sound just because they like it. When my friends from the UK visited me in Poland, they noticed the word “kurczaki” (pronounced [kurčak’i]) written on the shop window. The word means simply “chickens” in Polish, and it was used in its normal context as the shop was simply selling meat. My friends found the sound inexplicably funny and started to use it to refer to anything bizarre. Was this a case of meme transfer or the birth of a new meme? The answer depends on the identity conditions that we associate with memes. If we understand memes in purely behavioral terms, then the act of reproducing any action (word utterance amongst them) counts as meme transmission - no matter whether it retained its original meaning and cultural context. As I mentioned earlier, this understanding of memes could be useful, because it could explain the reasons why memes can be so easily used for misinformation. Nonetheless, once you look closely at what Dennett says about the transmission of words, it becomes clear that what he has in mind is probably not mere transmission of word-sounds or word-inscriptions.

Even though the book contains parts which suggest a very straightforward mechanism of meme transmission – for example, Dennett says that memes are reproduced perceptually (Dennett, 2017, p. 206) and that they are installed in our brains without being understood (Dennett, 2017, p. 213) – many other parts of From Bacteria to Bach and Back paint a much more complicated picture. At one point Dennett mentions that

> Children acquire the meanings of most of these words gradually, usually not even noticing the onset of comprehension, by a process that is not much like deliberate hypothesis formulation and testing, except in its capacity to secure results reliably: unconscious, involuntary statistical analysis of the multifarious stimuli they encounter (Dennett, 2017, p. 274).

This strongly suggests that the word copying procedure has to involve meaning. But if this is to be the case, then the explanation of the whole process is simply too sketchy. Dennett very
Can memes explain the birth of comprehension?

rarely talks about meanings which makes it difficult to ascribe any specific meaning theory to him. Still, there is no doubt that some theory is needed if we are to understand how memes are supposed to actually work. Naturally, Dennett does not deny the existence of word meanings. At one point he says that there are moments where we may have a feeling of having access to "bare meanings", typically in so-called "tip of the tongue" situations. (Dennett, 2017, p. 184). In another place, he speaks of the possibility of learning a word by listening to it in circumstances that give its meaning. How is this process of replicating word-meaning compounds supposed to look like? While it is easy to understand the "word" part as it consists in reproducing the sound so that it is close enough to the original, the "meaning" part demands further description. How is the user to copy the circumstances, context, etc.? I am not trying to say that the answer to this question cannot be successfully delivered. We could, for example, suggest a solution along the lines of directival theory of meaning (Grabarczyk, 2017) according to which users try to learn the meaning by remembering iconic, prototype circumstances in which the words were presented to them, but it is very hard to say whether this theory is something that Dennett would like to combine with his meme theory. This lack of details results in a huge detriment for memetics as it is practically impossible to determine the identity conditions for memes. Are they "the same" memes only if both the pronunciation and meaning remain close to the original? If so, how close do they have to be? Which of the two aspects is more crucial? To see this, consider the following three cases:

1) Polish word "idea" and the English word "idea". Both look identical and mean the same thing. Still, they are pronounced in a slightly different manner.

2) Polish word "aktualnie" and the English word "actually". They are similar in writing and pronunciation but mean two different things.

3) The English word "son" and the Ancient Greek word "ὑἱύς" (pronounced [hyː.ós]). Their meaning is the same and both originate from the same proto-Indo-European root, but they do not sound alike.

Which of these three pairs should we treat as two different memes and which as two exemplars of the same meme? The inability to answer questions of this type is a major sin in the case of a theory which introduces a new theoretical posit.

At one point, to explain the mechanics of meme copying, Dennett invokes a comparison with viruses:

An informational thing doesn't have a mind, of course, any more than a virus does, but, like a virus, it is designed (by evolution, mainly) to provoke and enhance its own replication, and every token it generates is one of its offspring. The set of tokens descended from an ancestor token form a type, which is thus like a species. (Dennett, 2017, p. 189)

He uses this analogy one more time when he suggests switching from our point of view to the memes' point of view (Dennett, 2017, p. 205). It is obvious that we should not read too much into this metaphor, because memes do not have any point of view in a literal sense. What Dennett means by this is that memes have an "agenda" – just like viruses, they try to stay in circulation as long as possible and as widely as possible. Moreover, we can be sometimes manipulated by memes just as our bodies are manipulated by viruses. They are installed in our
brains because of their attractiveness and not because of their beneficial nature. The trouble with this example is that in the case of viruses we understand both the mechanism of their replication and the reasons why they can infect the organism. We know exactly how they entice or trick our bodies. We do not have anything like that in the case of word-meaning compounds. Why are some of the words picked by the users and replicated quickly while others fade out from our vocabularies and conceptual repositories? Is this not the most important thing a theory of memes should explain?

It is quite probable that some of the properties of our behaviors make them more likely to be noticed, remembered and repeated. Being humorous seems to be a likely candidate for such a useful property. It is also quite probable that repeating some memes may be beneficial for us, at least in the short run. To use the example of humor once again – copying a “funny” meme may improve our social status. But this suggests that there are two sides or components of memes: one that tries to be preserved and one that exists solely to make the preserved part attractive to the carriers. How can we differentiate between them?

The analogy with viruses makes it all the more likely that the theory of memes could help us understand the dangers of their dissemination in controversial cases I mentioned before as it makes them even more similar to the so-called "dog whistles". And yet, as intriguing as this sounds, it is very hard to imagine how exactly we should proceed to flesh out this idea further. One simple example of why the development of the theory in this direction will be difficult is that words alone are rarely funny by themselves (as in my case of the word "kurczaki"). They are typically just parts of the bigger wholes that make them funny (or generally attractive). This leads to another question which is well-known from the debates surrounding the theories of meaning – how is the compositionality of memes supposed to work? Are we to assume that words try to be replicated by keeping the right company with other words just to be picked up as attractive sentences and stories?

To summarize, if words are to function as a model case for memes transmission, then the description of this case has to be: a) clear enough for us to understand the mechanism itself, b) universal enough to be extrapolated to other members of highly heterogeneous category of memes and c) able to give us at least a glimpse into how an agglomeration of competences can finally produce comprehension.

4. Where did comprehension come from?

So far, I focused mostly on the two former points. Unfortunately, the theory of memes does not fare much better when it comes to its role in the explanation of the transition from competence to comprehension. If the "installation" of memes in our brains does not require comprehension (and this is, after all, an important part of the theory), then the birth of comprehension seems to be even more mysterious. We collect memes, gaining more and more competences which – possibly through some additional interaction – give rise to comprehension. But why exactly does this happen? Is this just sheer luck or is it a necessary result of some unspecified

---

5 This example seems especially fitting since Dennett coauthored a whole book devoted to the phenomenon of humor (Hurley et al., 2017).
level of complexity of memes? Are there memes that could not be replicated by cognitive systems that are unable to comprehend? The answers that we get from *From Bacteria to Bach and Back* sometimes seem to be in conflict. At some point, Dennett mentions that in the case of some of the more complex information structures, their users had to master comprehension:

> What is particularly important in this exploration of memes is that some of these higher levels really do depend on comprehension, not just copying competence, even though they are based on, and rely on, systems of copying competence that do not require comprehension. (Dennett, 2017, p. 228).

In another place he declares that comprehension is never the source of competence (Dennett, 2017, p. 94). Are more complex theories that we install in our brains thanks to comprehension not to be treated as new competences?

It is worth adding that even in the case of cognitive processes which require comprehension in one cognitive system nothing precludes another cognitive system from achieving the same level of competence without comprehension. Dennett invokes the example of translation – a process which, in his opinion, requires comprehension – produced by a merely competent system, Google Translate. This leaves us with two crucial questions. Firstly, how does comprehension emerge from the agglomeration of competences?. Secondly, why did it appear in the first place? At some point (Dennett, 2017, p. 220) Dennett cites Mercier and Sperber (2011) and suggests that the birth of comprehension is somewhat related to our need for persuasion. If I read this suggestion correctly, it seems to present us with the following picture.

We collect and develop competences that help us successfully tackle the challenges of everyday life. These competences typically do not require comprehension. Since from the perspective of biological evolution comprehension is a fairly expensive function – it uses more energy and is probably also more time demanding than mere competence – we do not have access to most of these competences and perform them automatically. The example of the usage of words can be quite useful here. Our linguistic competence is something that we develop slowly through statistical analysis of contexts in which we encountered words as used by our peers. This experience allows us to use these words in similar contexts. This is how we "grasp" meanings. Since all of this can be achieved via mere competences, we do not have access to the meanings and are typically unable to list the conditions under which a given word can be used. And yet, at some point, because of the needs of social cooperation, we have to "explain ourselves" to others.

Needless to say, reframing the move from competence to comprehension as the move from action to the explanation of the reasons for this action does not automatically solve the mystery of comprehension. You might say that this solution only passes the question further as everything hinges on how exactly we have evolved the ability to give and understand reasons.

This capacity would not be possible without some sort of evolutionary foundation – competences that enable cognitive systems to read and signal their motives and goals without a full-blown comprehension of these actions. As pointed out by Zawidzki (2013), this is indeed the case. Evolution equipped humans not only with mind-reading skills but, more importantly, it produced many mechanisms that shaped their minds so they could become easier to read. The
ability to give reasons can be seen as the pinnacle of this long, multi-staged process. The upshot of understanding comprehension in terms of giving and requiring reasons is that this ability does not appear in the explanation out of thin air – we can actually follow and understand its complex development.

The idea that the pressure for better cooperation made competent cognitive agents master the art of “giving and asking for reasons” connects Dennet’s theory with Robert Brandom’s expressivism (Brandom 1998), something Dennett acknowledges, but never explores further. Similarly to Dennett, Brandom also starts with the description of human cognition as a statistically driven, automatic process which humans have to externalize to communicate and cooperate. We act in a non-reflexive, automatic way and for other people to understand us and predict our future actions we have to explain ourselves by providing reasons. It can be even argued that the distinction between ought-to-do and ought-to-be which Brandom borrows from Wilfrid Sellars could be mapped onto Dennett’s distinction between competence and comprehension respectively. The important difference between both theories comes from Dennett’s focus on the social function of persuasion. The aforementioned lack of access to our competences makes the process of communicating our reasons a difficult task. We cannot really explain our reasons as we do not really know them in the first place. The best we can do is that we try to reverse engineer our competences. The problem is that this process is just as difficult as the process of reverse-engineering the Alpha Zero. Additionally, since the main function of giving reasons is cooperation with others, we can sometimes achieve similar results by fabricating reasons. The most interesting aspect of this process is that we may sometimes be even unable to tell if what we do is the former or the latter.

The picture I presented above is interesting, but it does paint comprehension in a rather surprising way (provided that I interpret Dennett correctly). At best, comprehension seems to function as a completely independent cognitive function which may sometimes luckily align with competence it tries to make explicit. At worst, it looks like a facade.

Needless to say, the fact that this origin and function of comprehension may seem a little bit disappointing does not make Dennett’s analysis incorrect. What is problematic though is that once we accept this perspective, explaining the reasons why comprehension allows us to capture more complex memes – a claim Dennett is making – starts to be difficult to explain. It seems to me that it is this point of the theory where the underdeveloped connection with Brandom could be used fruitfully. A crucial aspect of Brandom’s account is the normative nature of language. It is not just that we are expected to present our peers with reasons for our actions – we are also expected to present the right reasons in the right situations. The fruits of our reverse-engineering efforts are evaluated by the community and this evaluation does not always favor our persuasiveness. This is why using logic and evidence differs from sheer rhetorical prowess. The combination of reverse engineering with the constraint of the norms society puts on reason-givers seems to me to be a more convincing story of how competences turn into comprehension. And we do not need to mention memes to tell it.
Acknowledgements

The work on the paper has been financed by the National Science Centre in Poland, thanks to the grant (2014/15/B/HS1/01928).

Bibliography


The editorial and publishing process of this publication has been financed by the Ministry of Science and Higher Education from the funds for the dissemination of research (DUN) within the framework of publishing activity, contract no. 711/P-DUN/2019, period of implementation: the years 2019–2020.