Let's Talk AI with Onur Güntürkün

Onur Güntürkün[1] and Barbara Steffen[2]

[1] Ruhr University Bochum, Department of Biopsychology, Germany,
onur.guentuerkuen@ruhr-uni-bochum.de
[2] METAFrame Technologies GmbH,
barbara.steffen@metaframe.de

"AI gives us the unprecedented chance to investigate the commonalities and differences between AI and BI (biological intelligence)."

The Interviewee - Onur Güntürkün

My Personal AI Mission:
I'm interested to what extent we can learn from AI about BI.

My Takes on AI

Artificial Intelligence: A non-biological system that can learn from experiences, generates past encounters to new and different encounters and tasks, solves problems, reasons about causes and consequences, is able to adapt to new situations, understands and handles abstract concepts, steers its body through complex novel environments and successfully manipulates its environment in endless unforeseen situations with its limbs. According to this definition, AI is still in its infancy.

Trust: I never fully trust another intelligence.

Explainability: I'm a cognitive neuroscientist. My task is to run studies to explain BI. An AI that cannot be explained can be a tool, but nothing that interests me in scientific terms.

Essential Elements of Human Capabilities: Look at my definition of artificial intelligence.

The Interview

Barbara *Today, I have the pleasure of speaking with Professor Onur Güntürkün from Ruhr-University Bochum. Could you briefly introduce yourself and your connection to artificial intelligence?*

Onur I'm Onur Güntürkün, a biological psychologist at Ruhr-University Bochum. I identify myself as a comparative cognitive neuroscientist. I work with a variety of animals, including humans, pigeons, crows, Nile crocodiles, dolphins, and many more. As a cognitive neuroscientist, I'm interested in thinking, memory, and conclusions. As a neuroscientist, I study the brain, specifically the link between brain and cognitive function. My relationship with AI is that I work on BI, or biological intelligence. AI is a parallel avenue of research, and it's not easy to see the familial relationships between BI and AI. While AI and BI may seem similar in their effects, especially in recent years, a deeper look reveals that the similarity is more superficial and not necessarily related to the mechanisms. This is a fascinating and intriguing area of research.

Barbara *Are there specific research questions you're currently addressing in the context of artificial intelligence, or is it rather a general interest?*

Onur Maybe there is one more specific question. The large language models developed in recent years pose intriguing questions to biological intelligence. If you ask cognitive scientists about cognition and its constitution, they might argue that two main areas of cognition exist. The first is based on associations. For example, we understand the relationship between cutlery and a dish on the table, or a bottle and beer, because they occur in close association. This associative account allows us to predict the future, much like what large language models do with words [7]. Then there is a second area of cognition, which involves drawing conclusions, inferring reasons, understanding causality, and more. Many cognitive scientists assume that this second branch cannot be based on "simple" associations [3]. But AI starts to challenge this view [6].

"Biological intelligence is now realizing that AI is suddenly addressing core questions on the structure of cognition, and it's possible that the tables are turning, with AI contributing more to biological intelligence research than before."

Barbara *How do we form these associations? Is it through observation? Do we all see similar associations or patterns, or is it unique to individuals?*

Onur It's the stream of life [8]. For instance, I'm looking at your face, and I know your name is Barbara. I'm at a university, I know I look into my monitor. All these stimuli are now associated with each other. If I see you in a different location, I might not recognize you immediately because you're out of context. But upon closer inspection, I would recognize you as Barbara and start a conversation about artificial intelligence, as that's the associative cloud you're embedded in. Our brain is constantly learning these associations, and this forms

the majority of our cognitive processes. Some cognitive scientists argue there's more to it that is not based on pure association, and that's the discussion we're having now.

Barbara *Do you think that the development and progress of artificial intelligence is helping us to question and look at biological intelligence differently? Does it help us derive new ideas and questions that help us make progress in established fields like biological intelligence?*

Onur Absolutely. Until now, biological intelligence has nurtured artificial intelligence. AI was inspired by neuroscience, and while AI still draws inspiration, it's doing its own thing. Biological intelligence is now realizing that AI is suddenly addressing core questions on the structure of cognition, and it's possible that the tables are turning, with AI contributing more to biological intelligence research than before [5]. There might be a shift in this interdependency.

Barbara *How do you view the role of trust in the adoption of AI?*

Onur That's a complex question. Google started with "Don't be evil," and now we see where Google is. It's impossible to say there are things we all agree not to do because someone else will do it. During the Cold War, the reason we didn't have a Third World War was that both sides could inflict unimaginable damage on each other. We might be heading towards a world where AI becomes so powerful that either democratic systems curtail the possibilities of large companies or nations, or there's a balance of threats that limits the potential damage. I'm just speculating here, and I don't know if this will happen.

Barbara *Do you think we are currently overestimating or underestimating the power of artificial intelligence?*

Onur I'm not sure. We seem to oscillate between the two. There are still people who argue that AI will never achieve certain cognitive aptitudes. But we've been saying this for the last 30 years. They said AI would never beat a chess player, become a Go champion, and so on. All of that turned out to be wrong [4]. I don't believe in setting limits on what AI can do. Even the Moravec paradox, which suggests that robots can easily do what's difficult for us but struggle with what's easy for us, might one day be proven wrong. I don't see any reason to limit AI yet.

Barbara *You've studied many animals and their intelligence. Through various experiments, you've discovered that animals have cognitive traits that we once thought were unique to humans. Do we, as humans, tend to overestimate our superiority and set up specific criteria to demonstrate our uniqueness? Do we try to prove that we are superior to other animals and machines?*

Onur The central pursuit of science for the last 400 years has been to demonstrate that we are not alone, that we are founded on the same mechanisms as other living beings, and that the universe does not revolve around the earth, but rather, we revolve around the sun, which is a very average star. These lessons are results of a long pursuit from the natural sciences. I contribute a bit to this by

demonstrating that the cognitive repertoire of many animals, even those considered as lowly, overlaps significantly with the basic principles our own cognition. Furthermore, different animals develop the same brain mechanisms to produce very similar cognitive operations, suggesting that our brains are much more alike, at least in their functional architecture, than we previously thought. It appears as if nature has severe limitations in the degrees of freedom it has to create intelligent organisms, leading us to increasingly refer to similar mechanisms [2].

Barbara *When you think about concepts like trust, do you think we often rely on instinct, or do we consciously and cognitively decide when and in what context to trust someone, a machine, or an application?*

Onur As social animals, we need to trust other individuals, which is fundamental to being part of a social group. However, living in a social group doesn't mean we always trust others. Many animals live in groups solely because their survival rate is higher in a group than when alone, but that doesn't mean they trust their neighbors. For instance, I often work with pigeons. Pigeons live in flocks to reduce the risk of being killed by predators, but they don't care about their neighbors and don't interact with them in a cooperative sense. Humans and several other animals are different in this regard. We interact a lot and care about the well-being of our neighbors. But this trait obviously opens up the possibility for exploitation. The human brain has developed interesting mechanisms for trusting or distrusting others, with individual differences. These are all part of the DNA of Homo sapiens. It's fascinating how easily these trust systems can be fooled by simple mechanisms. So, can I trust a human being to be trustworthy? I can't. I grant trust to a person based on my past experiences. The game with AI is the same, just with something that isn't alive.

"I don't believe in setting limits on what AI can do. Even the Moravec paradox, which suggests that robots can easily do what's difficult for us but struggle with what's easy for us, might one day be proven wrong."

Barbara *When two people interact, they both have their own intentions, emotions, and feelings. I cannot simply respond to your needs and feelings because I am bound by my own. In contrast, artificial intelligence is very good at mimicking or acting in a way that it believes the human wants or needs in that situation. Do you think that interactions with machines are riskier because machines are not tied to their own emotions and feelings and basically have no direct long-term consequences?*

Onur At its core, there might not be much difference because you can train a person to appear trustworthy without being so. However, an artificial system can converse simultaneously with 100,000 people. So, the difference isn't in the principle, but in the system's ability of widespread use.

Barbara *Do you have any specific measures in mind that we should build into AI systems to ensure ethical adoption of AI?*

Onur I wish there were, but that's not my field. I'm just observing and hoping for the best.

Barbara *Looking into the future and the potential capabilities of artificial intelligence, on a scale of 1 to 10, where 1 represents the AI systems we know today like ChatGPT and Gemini, and 10 represents something like artificial general intelligence. What do you think will be possible?*

Onur 10. Definitely, 10.

Barbara *Could you elaborate on that?*

Onur Considering the rapid pace of development and extrapolating from that, I see no reason why growth should be limited anytime soon. Therefore, I would rate it a 10.

Barbara *How long do you think it will take for an AI system to reach artificial general intelligence?*

Onur It depends on what you mean by intelligence. If you're referring to the ability to communicate, solve equations, and so on, then such systems already exist. If you're including more complex tasks like moving, planning, and executing various activities, that will obviously take longer, but I'm confident it will happen. I can't provide a specific timeline, but I'm certain it will occur. I don't subscribe to dystopian views of machines taking over, but I do believe that these machines can be exploited by humans and other machines, potentially leading to dystopian outcomes. However, I'm uncertain about where it will end.

Barbara *What are essential human capabilities that are currently difficult for machines, but could be possible in the future?*

Onur Well, computers have already beaten chess and Go champions, but a human still has to move the pieces. This suggests that physical movement is the biggest hurdle. So, I would say that's an area where improvement is necessary for general intelligence. I'm not sure how long it will take. For the rest, we're increasingly retreating into smaller and smaller niches of argumentation. We once thought that drawing conclusions and making mathematical proofs were impossible for AI, but it turns out they are achievable. So, there's not much left. Then, of course, we can retreat into things like, "but they don't feel," and so on. But that's a cheap retreat. As I said, I gave them a 10, believing that one day they will perform as we do.

"Different animals develop the same brain mechanisms to produce very similar cognitive operations, suggesting that our brains are much more alike, at least in their functional architecture, than we previously thought."

Barbara *Reflecting on our future with AI on a scale from utopia to dystopia. Where do you stand?*

Onur I don't really believe in dystopian concepts like singularity, but I do have dystopian fears about artificial systems being used by humans to make life miserable for others. At this point, I think anything is possible, especially since we're not entirely sure what could happen. We might be living in the most dangerous period because artificial intelligence is developing rapidly, and we're ill-prepared for its potential implications. In 20 years, we may have experienced enough to know where the threats lie and how different people might use AI against others, giving us more defensive options. Currently, we're living in a dangerous utopia, believing we're safe when we're not.

Barbara *We're not safe because there are always people with bad intentions?*

Onur Yes, there will always be people with bad intentions. Remember the first computer viruses? People were completely unprepared for them, and it was easy for a simple computer virus to cause damage. Once viruses became more widespread and people realized the threat, defense mechanisms were developed, leading to an arms race between virus programmers and antivirus software developers. We're in a similar period now, where we're not fully aware of the potential dangers of AI. A villain with a bright mind could exploit something that, in hindsight, seems simple, but we're not prepared for. That's what I mean when I say we're in a dangerous period where we don't know exactly what could happen.

"We might be living in the most dangerous period because artificial intelligence is developing rapidly, and we're ill-prepared for its potential implications."

Barbara *Do you think we might lose control as we continue to increase computing power and data, leading to increasingly advanced AI systems with unforeseen emergent properties?*

Onur I could imagine that. Remember HAL 9000 in 2001: A Space Odyssey [1]? HAL 9000 wasn't evil; it was programmed to serve the mission. It concluded that the astronauts were jeopardizing the mission, so it had to choose between not lying to humans and serving the mission. It decided that serving the mission was more important than not lying to humans. This could be a glimpse into our future, as the complexity of these systems is so vast that we can't predict what conclusions they might draw in conflicting situations. We're living in interesting times.

Barbara *It's scary, isn't it?*

Onur Everything is scary, but scary things are also interesting.

Barbara *That's true. So, should we be careful about the mission we give AI systems to ensure that they don't disregard certain principles, such as not lying to humans, in their pursuit of the mission?*

Onur It's hard to predict. The programmers of HAL 9000 didn't have bad intentions when they programmed it to prioritize the mission. They couldn't foresee the consequences. This could happen at any time.

Barbara *It's interesting because it's normal for us to put people in positions of power or responsibility to make tough decisions. But now that we are faced with the possibility of delegating that responsibility to an AI system, we are starting to wonder if that is the right thing to do. Do you think the problem is that it is a system or that we are becoming dependent on it on a much larger scale with a much greater potential for harm than if a human were to make a bad decision in a particular situation?*

Onur That's possible, but don't put too much trust into logical decisions of humans. Psychology is full of examples where human decisions change based on minute details. For example, in the classic trolley problem, you have to decide between five people being killed by a trolley that is out of control, or you pull a lever to divert it to another track where it kills one person. Most people pull the lever. But if you must physically push one person on the track to stop the trolley from killing five people, you're less likely to do it, even if it means saving five people. This doesn't make sense, but that's how we think. Whether artificial systems would make more prudent decisions, I don't know.

Barbara *Fascinating. What if we rely more and more on artificial intelligence? I've heard that the first universities are considering stopping bachelor theses because it's becoming difficult to tell whether the AI or the student is doing the work. Do you think there's a risk that as we outsource more and more of the skills we've relied on in the past, we'll lose those skills in the future?*

Onur Well, as an experimental scientist, I must say that our bachelor, master, and PhD theses are experimental, and AI cannot replace that. So, my life is easier. I don't mind if my students use ChatGPT to structure their discussions or write parts of their discussions or introductions. It's more challenging for scholars in the humanities or text-based academic fields to cope with AI in teaching. This will certainly change their structure and approach to scientific research. I'm unsure about the direction it will take.

Barbara *There are those who argue that writing is a way of thinking and structuring thoughts. It's like having a self-discussion with a piece of paper. If you just type in a few prompts and wait for ChatGPT to create a sufficiently logical and convincing argument, could we lose understanding?*

Onur Yes, this might happen. As a university teacher, I could see that individual oral exams could solve this problem. The student sits with me for 30 minutes, and we discuss things. This could re-emerge because the written part might not be enough. I can read the thesis, but when I talk to the student about the thesis and realize they don't understand their own written arguments, it's a litmus test. This could be a re-emerging trend. Barbara, you're asking a lot about the future. I really don't know.

Barbara *Yes, I understand. But it's now that we have the choice to either "wait and see" or to think about what we want, what is acceptable to us, and what we can do to nudge the future trajectory toward the desired future. So how can we nudge it in a direction that seems favorable at the moment?*

Onur I agree entirely. In my discipline, talking to a student after they submit their thesis and discussing everything related to it could be a solution. Even if you read what ChatGPT wrote or created, there's a difference in understanding the text once you start discussing it with your professor. This could be a way out.

Barbara *Absolutely. Are there specific areas where you'd like to see more interdisciplinary collaboration, especially now that AI is on the rise?*

Onur I'm already deeply involved in interdisciplinary work, so I value it. Regarding AI, the developments in AI and our understanding of cognitive science as an experimental academic field, I believe we need more interaction at that level.

Barbara *Are you also referring to finding ways to limit the current pace of progress being driven by organizations? Making sure that we address and integrate legal and ethical implications and think about the potential impact of AI tools on humans from a psychological standpoint. Identifying a structure and framework that seems more favorable than just letting big tech organizations come up with new ideas for automation and short-term profit.*

Onur Yes, certainly. Despite its name, Open AI is not as open as it suggests. We should use all ways of communicating with each other without relying solely on legal avenues. I'm happy to talk and interact with anyone. That's my academic freedom, and we should use it. As I said, we now have two black boxes. One is our mind, and the other are large language systems. Uri Burda and Harry Edwards from Open AI admitted they have no idea what's happening in the system. So, by joining forces, we might be able to understand the common and separate elements of these black boxes.

Barbara *What should be our vision for AI and its future development in the coming years?*

Onur AI can be developed in tech companies for applications without our involvement. If we want to understand it at an academic level, to truly comprehend what we call cognition, then we need to foster more collaborations between cognitive scientists and AI experts. I thought we had enough of that, but I now realize we probably don't. We have theoretical neuroscientists on campus, and they're my close colleagues. They're closer to what, for example, OpenAI is doing, but they're still too far away. The problem is that what's done at the university level simply doesn't have the financial and organizational means to develop what tech companies are doing. That's where interdisciplinarity stalls, at least in my view.

Barbara *It's interesting that we see such great progress, but I always wonder how much of it is conscious progress and how much of it is experimental progress.*

People with money and power are trying things that may lead us into a future where we face consequences that we could have anticipated if we had progressed at a slower, safer rate for humanity and all life on this planet.

Onur Barbara, I'm sure that as we speak, developments are happening in certain companies that we might later wish had taken a different path if they had consulted with person X or Y. If this would happen at university levels, it would be easy to implement measures because we've been doing interdisciplinary work for decades. But there's a real gap between large tech companies with billions of dollars at their disposal and universities. Universities are interested in money, but they're also interested in science and publishing in prestigious journals. Their values and rewards are different. Tech companies, understandably, have to be restrictive in their communication because they have to make money. These differing values cause communication problems and often make it one-sided.

"If we want to truly comprehend what we call cognition, then we need to foster more collaborations between cognitive scientists and AI experts. I thought we had enough of that, but I now realize we probably don't."

Barbara *That's true. It's somewhat alarming that we see this arms race where every organization wants to be first, which incentivizes them to take shortcuts that ensure significant leaps to stay ahead of the competition. Is there anything else you would like to add to this interview?*

Onur No.

Barbara *Onur, thank you very much for your time and insights. I especially enjoyed learning more about the perspectives of biological intelligence on artificial intelligence and vice versa! Have a nice day.*

References

1. Clarke, A. C., 2001 – A Space Odyssey (Hutchinson, 1968).
2. Güntürkün, O., Pusch, R., Rose, J., Why birds are smart, Trends in Cognitive Science, 2024, 28: 197-209.
3. Kahneman, D. Thinking, fast and slow (Macmillan, 2011).
4. Kasparov, G, Deep Thinking (John Murray, 2017).
5. Kotseruba, I., Tsotsos, J. K., 40 years of cognitive architectures: core cognitive abilities and practical applications, Artificial Intelligence Review, 2020, 53:17–94.
6. Moskovitz, T., Miller, K., Sahani, M., Botvinick, M. M., A Unified Theory of Dual-Process Control, arXiv:2211.07036v3 [q-bio.NC].
7. Sherman, B. E., Graves, K. N., Turk-Browne, N. B., The prevalence and importance of statistical learning in human cognition and behavior, Current Opinion in Behavioral Sciences 2020, 32:15-20.
8. Vong, W. K., Wang, W., Orhan, A. E., Lake, B. M., Grounded language acquisition through the eyes and ears of a single child, Science, 2024, 383: 504–511.