David Gamez

Which Properties and Structures must a Virtual World have to Enable the Successful Creation of Human-level Artificial Intelligence?

david@davidgamez.eu

Thursday, March 30, 2023

Which Properties and Structures must a Virtual World have to Enable the Successful Creation of Human-level Artificial Intelligence?

Introduction

Discussions about intelligence often distinguish between the *type* and *level* of intelligence. Intelligence *level* refers to how much intelligence a system has. Within psychology, intelligence level is often measured using IQ and *g*: someone with an IQ of 130 has more intelligence (a greater level) than a person with an IQ of 90. IQ and *g* are population-based measures that are difficult to apply to diverse non-human animals and AIs. This has led to the development of absolute (ratio scale) universal measures of intelligence that can, in theory, be applied to any system at all – see, for example, Legg and Hutter (2007), Hernández-Orallo and Dowe (2010) and Gamez (2021).

As Section 1 explains, there are many distinct *types* of intelligence that specialize in different environments. AlphaGo has a high level of intelligence in the game of Go, and almost no intelligence outside this environment. Humans have high levels of intelligence in the three-dimensional human-scale natural world. They have less intelligence than AlphaGo within Go, and almost no intelligence in high-dimensional data-rich environments.

In the title question, human-*level* AI presumably refers to an AI that has the same *amount* of intelligence as a human. However, it is not specified whether this AI should have the same *type* of intelligence as humans. It is conceivable that a human and an AI could have the same *amount* or *level* of intelligence (assessed using a universal measure) without their intelligences being the same *type*. One could be intelligent about physical properties of the natural world; the other could specialize in high-dimensional relationships between gravity waves.

The type/level distinction is important when we are designing virtual environments for the creation of artificial intelligence. If we want an AI whose intelligence is human-like *and* human-level, our virtual worlds will almost certainly have to reproduce the natural environments in which human intelligence flourishes. The AI will have to be embedded in a virtual body whose mechanical properties and senses are similar to a real human body - the ECCE robot is a nice example (Wittmeier et al. 2011). It would be helpful if there was social transmission of culture.

Human-like intelligence is extremely useful in the world that humans inhabit. AIs with human levels of human-like intelligence will have many applications. However, human-like intelligence has significant limitations – for example, we cannot reason about large data sets or mentally manipulate five-dimensional objects. Human-like intelligence might not be the best kind of intelligence for solving difficult problems, such as climate change.

The first part of this essay explains why intelligence is not an absolute property of a system. A system's intelligence has different levels in different environments and a limited ability to generalize beyond the environments that it is specialized for. If we want to develop non-human forms of intelligence that are potentially better at solving global challenges, then we might need to design virtual worlds that lead to AIs with forms of intelligence that are more adapted to solving these problems. Some ideas for this are set out in Section 2.

Natural systems perceive their physical environments through their senses and have elaborate ways of processing raw sensory data into experienced worlds, which are often referred to as *umwelts* (Uexküll 2010). Different systems inhabit different umwelts, and intelligence develops and operates inside umwelts. Section 3 discusses the relationship between umwelts and intelligence and suggests how different combinations of virtual worlds, virtual bodies and virtual senses could lead to different types of intelligence.

1. The Generality of Intelligence

Intelligence is often believed to be an absolute property that is independent of a system's environment. However, real human intelligence varies considerably with the environment. Someone who has played video games all their life has a much higher level of intelligence inside video game environments than members of Amish communities, who have never used computers. On the other hand, Amish people are likely to have much higher levels of intelligence in agricultural environments. All humans have low levels of intelligence in hundred-dimensional environments and in environments containing petabytes of numerical data. As Chollet (2019) points out, the human brain evolved to help us survive in a hunter-gatherer environment and it has a limited ability to generalize beyond this environment:

> We argue that human cognition follows strictly the same pattern as human physical capabilities: both emerged as evolutionary solutions to specific problems in specific environments (commonly known as "the four Fs"). Both were, importantly, optimized for adaptability, and as a result they turn out to be applicable for a surprisingly greater range of tasks and environments beyond those that guided their evolution (e.g. piano-playing, solving linear algebra problems, or swimming across the Channel) – a remarkable fact that should be of the utmost interest to anyone interested in engineering broad or general-purpose abilities of any kind. Both are multi-dimensional concepts that can be modeled as a hierarchy of broad abilities leading up to a "general" factor at the top. And crucially, both are still ultimately highly specialized (which should be unsurprising given the context of their development): much like human bodies are unfit for the quasi-totality of the universe by volume, human intellect is not adapted for the large majority of conceivable tasks. This includes obvious categories of problems such as those requiring long-term planning beyond a few years, or requiring large working memory (e.g. multiplying 10-digit numbers).
>
> Chollet (2019, pp. 22-23)

IQ tests and the work on *g* have led some people to believe that intelligence is a completely general-purpose ability. A person's IQ is thought to correspond to their level of intelligence independently of the environment they are in. However, a person's IQ only corresponds to their intelligence within academic environments that are similar to IQ tests. This is why IQ is often justified by pointing to correlations between IQ score and success in professions that require academic skills (Robertson et al. 2010). A person's IQ does not correspond to their level of intelligence in non-academic environments – all people have zero intelligence in hundred-dimensional environments or in environments constructed from large numerical data sets.

Ideas about artificial general intelligence (AGI) are usually derived from beliefs about the generality of human intelligence. If human intelligence is not completely general, then there is very little reason to believe that completely general artificial intelligence is possible. A much more plausible view is that there are many different types of natural and artificial intelligence that are optimized for different environments. This idea has often been discussed in the literature on intelligence. For example, Gardner (2006) claims that there are multiple types of intelligence, including musical intelligence, linguistic intelligence and emotional intelligence. Warwick (2000) frames this more generally with his idea that intelligence is a high-dimensional space of abilities.

If intelligence is, to a greater or lesser extent, specialized for one or more environments, then a system's level of intelligence should be indexed to the set of environments in which it has that level of intelligence. I have one level of intelligence in *Minecraft*, another level of intelligence in poultry farming, and zero intelligence in high dimensional spaces populated with genomics data.

2. Customizing Virtual Worlds for the Problems We want to Solve

Human-level human-like intelligence is extremely useful in three-dimensional environments that are similar to the natural world. For example, Shanahan (2015) describes how teams of virtual human-like and human-level AIs could solve problems much more rapidly than real humans. However, human-like AIs often have the wrong kind of thinking to address challenging problems. It is for this reason that many AIs are not optimized for the human world at all – they are designed to identify patterns in large quantities of genetic data, personal data for marketing, etc. These AIs are not human-like, but they exceed human levels of intelligence in the environments they are designed to work in.

To build AIs with human or superhuman levels of intelligence that can solve global challenges, we might have to design environments that are specific for those challenges, instead of copying the physical world and hoping that this will lead to a general intelligence that can solve any problem. These virtual worlds are likely to have much higher dimensions than the physical world. and they could be populated with datasets and scientific papers as well as representations of the physical aspects of the problem.

Radical transformations of the problem spaces might be necessary to develop intelligences that can solve the problems more easily. A nice example of this approach is the kernel trick that is used with support vector machines. This enables a collection of points that are not linearly separable to be mapped into a space in which they can easily be divided into two classes. Kernel tricks could be used to create virtual worlds in which AIs can solve specific problems much more easily than humans. This larger space of possible virtual worlds could enable us to create AIs with superhuman levels of intelligence that are highly optimized for the problems we want to solve.

3. Environments and Umwelts

Natural systems do not perceive all of their physical environments at once. They are embodied and perceive their local environments through their senses. Most AIs work in a similar way. Self-driving cars perceive their local environments through cameras and ultrasonic sensors; neural networks are trained by sequentially presenting data to their inputs. In the real world a system's access to its environment is mostly fixed by physical constraints on natural and artificial senses. This is not true in virtual worlds where, for example, it is possible to see for a thousand virtual miles or hear a virtual pin drop on another virtual planet.

Agents that cannot access their environments cannot develop intelligence in their environments. Agents with complete spatial and temporal access to a deterministic environment can predict what will happen under any conditions. The latter agents will have low levels of intelligence in non-deterministic environments and in environments that they cannot fully access. Intelligence typically develops between these extremes. There are important trade-offs between an agent's level of access and its generalization ability: greater access to the spatial and temporal environment increases learning speed and reduces generalization.

The design of a system's senses is also important for its intelligence. A virtual agent could sense everything within the spatial and temporal area that it has access to, or, like animals, its senses could be limited to specific aspects of its environment – certain frequencies of light and sound, a few chemicals, and so on. Sense design also involves trade-offs between computational resources, learning speed and generalization.

Natural systems have hard-wired and learnt ways of processing raw sensory data into abstract concepts that are tailored to their forms of life. For example, humans do not directly perceive the cascades of spikes generated by their photoreceptors: they see a world of colours, shapes and objects that are arranged in three-dimensional space. Self-driving cars use deep neural networks to classify raw sensory data as people, road signs, cars, etc. To reach human-level intelligence AIs are going to need *classifiers* that convert raw sensory data into representations that are relevant to the problems they are attempting to solve. These can be hard-wired, or the AIs can be given the ability to develop their own classifiers.

Spatial and temporal limits on perception, sensory filters and classifiers result in agents whose perceived environments are radically different from the virtual worlds they are immersed in. An animal's experienced environment is often referred to as its *umwelt* (Uexküll 2010), and there has been a lot of research on animal umwelts (Yong 2022). The umwelt concept is also essential for understanding AI behaviour. All intelligence develops and operates within the umwelts of natural and artificial systems - it is not applied to the physical world conceived from an objective perspective. For example, chemists predict that hydrogen and oxygen mixtures will explode when they are exposed to heat. Their intelligence is operating in a space in which the concepts of hydrogen, oxygen, heat and explosion have been abstracted out of the quantum flux of elementary particles. Non-human animals do not have these concepts and cannot make intelligent inferences about hydrogen, oxygen, heat and explosions.

The question about the design of virtual worlds for the creation of intelligence is the wrong question. We should really be asking how we can design umwelts for the development of intelligence. A virtual world can contain many different umwelts with diverse potentials for the creation of intelligence. Virtual worlds are sources of signals that cause changes in umwelts and can lead to the development of intelligence inside those umwelts. Changes in a virtual world that do *not* affect a system's umwelt will not affect the intelligence of that system. For example, humans have a very weak understanding of gravitational waves. They can only be intelligently understood by humans through specialized measuring instruments and socially transmitted knowledge.

In Section 1 I argued that human-like intelligence only reaches human-levels within environments that are similar to the hunter-gather environments in which human intelligence evolved. It is more accurate to say that human-like intelligence only reaches human levels inside umwelts that are created by the interaction between human-bodies and physical hunter-gatherer environments. If we want human levels of human-like intelligence, we will have to create virtual environments and virtual bodies, whose interactions result in human-like umwelts, inside which human-like intelligence can flourish. Human-level AIs with radically different forms of intelligence could develop within umwelts that are designed around specific problems we want to solve.

Conclusion

This essay has argued that a system's level of intelligence varies with the environment that it is in. Even human intelligence only reaches high levels in environments that are similar to the hunter-gatherer world in which humans evolved. Human-like intelligence with human levels can be created in virtual worlds that are similar to our natural world. We could also create specialized high-dimensional data-rich environments, in which AIs could reach high levels of non-human forms of intelligence. These new types of intelligence might be able to solve problems that are difficult or impossible for humans.

The natural and artificial systems that we have encountered and created so far have not had god-like access to their entire environment. Instead, they are progressively exposed to their environments, and data from their environments is often filtered through senses and processed into higher level abstractions through classifiers. The environment that each agent experiences is not the actual physical or virtual environment, but an umwelt in which the processed output from the senses appears. Intelligence develops and operates within umwelts and many different umwelts can coexist in a single physical or virtual world. If we want to create human-level intelligence, we should focus on the properties and structures of umwelts, not on the properties and structures of virtual worlds.

References

Chollet, F. (2019) On the measure of intelligence. *arXiv*: 1911.01547.

Gamez, D. (2021). Measuring Intelligence in Natural and Artificial Systems. *Journal of Artificial Intelligence and Consciousness* 8(2): 285-302.

Gardner, H. (2006). *Multiple Intelligences: New Horizons.* New York: Basic Books.

Hernández-Orallo, J. and Dowe, D. L. (2010). Measuring universal intelligence: Towards an anytime intelligence test. *Artificial Intelligence* 174: 1508-39.

Legg, S. and Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and Machines* 17: 391-444.

Robertson, K. F., Smeets, S., Lubinski, D. and Benbow, C. P. (2010). Beyond the Threshold Hypothesis: Even Among the Gifted and Top Math/Science Graduate Students, Cognitive Abilities, Vocational Interests, and Lifestyle Preferences Matter for Career Choice, Performance, and Persistence. *Current Directions in Psychological Science* 19(6): 346-51.

Shanahan, M. (2015). *The Technological Singularity.* Cambridge, Massachusetts: The MIT Press.

Uexküll, J. v. (2010). *A Foray into the Worlds of Animals and Humans, with A Theory of Meaning.* Translated by J. D. O'Neil. Minneapolis: University of Minnesota Press.

Warwick, K. (2000). *QI: The Quest for Intelligence.* London: Piatkus.

Wittmeier, S., Jäntsch, M., Dalamagkidis, K. and Knoll, A. (2011). Physics-based modeling of an anthropomimetic robot. 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, CA, USA: 4148-53.

Yong, E. (2022). *An Immense World: How Animal Senses Reveal the Hidden Realms Around Us.* London: Vintage Digital.