Aug. 20, 2020

Excerpt from a draft of a
forthcoming book on Intelligence
BY

Kristinn R. Thórisson

Made available to graduate students in the course

Advanced Topics in Artificial Intelligence
Given by K.R.Thórisson
In the fall of 2020 at Reykjavik University

1 Earth Offers Great Variety

Anywhere and everywhere we look in the cosmos, we see bleaker, poorer, blander worlds than what appears before us even by simply looking out the window. If you live in a city you may see houses with windows, doors, and roofs; if you live in the countryside your view may include mammals grazing, birds flying above, a lake, a river and ten types of plants. These phenomena have been all but ruled out on the many other planets we know about. Earth presents us with an estimated 100,000 types of trees, rocks of infinite shapes and textures, and water – the most abundant chemical bond on the planet – appears in many ontologically separate forms: oceans, lakes, rivers, waterfalls, streams, swamps, dams, rain, hail, snow, ice, soda pop, vapor, dampness, clouds, fog. If we happened to find two rocks that seemed very much identical in every way to our naked eye, we need only go down to the microscopic level to see that these hypothetically identical rocks are in fact not identical at all. Due to twists and turns of branches, none of the 3 billion trees on our planet are identical at the branch or leaf level of detail; material makeup, fractures, and indents and scratches set every single rock on the planet apart from everything else; every arrangement of cups, plates and cutlery in all the dishwashers in the world – since their invention – is unlike any other. Even if time were to stand still from now on and nothing ever changed, the world would harbor enormous complexity; nothing is exactly the same as anything else. The world is non-repeating at most scales that matter to our everyday existence.

It is not simply the seemingly static variety of natural phenomena that we encounter on Earth, the enormous variety of phenomena found display complex dynamics—variation in time. Everyday things change position at numerous time scales: From seconds to minutes to hours to days to weeks to months to years to decades. Over the span of 10 minutes walking around outdoors we may have encountered several things that occupy each of those levels of change: a fly (milliseconds), a flying bird (seconds), a crowd at a bus stop (minutes), shadows of houses on the street cast by the sun (hours), billboards (weeks), trees shedding their leaves (months), store signs (years), traffic signs (decades). Everything we see in our surroundings every day can be broken down into a large number of variables, each of which is subject of forces that may change their values from instance to instance and moment to moment, increasing the already great structural variety enormously: The size, general direction, and speed of raindrops makes today's rain shower look different from yesterday's; the movement of traffic today is not identical to its movements yesterday; vegetation, the seasons, urban sprawl, wind patterns, the dividing cells in your body—time is another obvious source of great variety in the world.

And yet there is another even bigger source of complexity—interaction. All entities in our everyday experience have a tendency to interact, in a multitude of ways, with a large number of other entities, creating emergent patterns that would not exist otherwise: The tires of vehicles wearing down the pavement; the river at the bottom of the valley making a snake-like path that changes slowly over years and decades; the temperature in the atmosphere freezing water drops in free fall, each containing 10^{18} water molecules, turning them into snowflakes with crystalline patterns, forming a microscopically unique pattern, then thawing them back; humans harvesting natural resources to generate energy for building houses, laying roads and digging tunnels; fires burning down whole villages;

cupboards holding cups and plates; cups breaking when they fall to the floor; your foot hitting a chair's leg when walking past it—the list making today different from yesterday is endless. The number of *relations* between entities in the world is vastly larger than the number of entities and their possible states combined, by allowing infinite combinatorics, making no encountered situation exactly alike—even if we look only at the macroscopic level, things we can see, feel and touch.

The variety keeps increasing as we take a detailed look at complex systems such as plants and animals. No one knows how life came about, but as long as 4 billion years ago the first life may have appeared, probably relying on sunlight and oxidation to extract energy from the environment. All living systems, including all estimated 5 billion species that have ever lived, need a regular supply of energy to maintain their processes; in the words of Ilya Prigogine, they are *dissipative structures* that need a regular influx of energy and removal of waste to maintain the function of all their processes. They are *self-sustaining* on Earth in that they are stable structures amid potentially destructive forces, came into being through natural processes, and maintain themselves through self-organizing natural processes.

Energy for animals comes in the form of plants and other animals; plants grow in different areas from year to year and decade to decade, animals move around all the time, many of them fleeing much faster than any humanoid could run. And some animals pose threats. Environmental variety means that every individual of a species will encounter a number of things in its environment that could not have been anticipated beforehand, could not be enumerated or hardcoded in genes. To handle it individuals must have the ability to classify natural phenomena, preferably in a way that allows the emerging classification to predict how to classify future phenomena, as in for instance the color and shape of a poisonous berry helps identify other potentially poisonous berries in a related tree species. Threats and food – the two major determinants of whether a species can survive – display a vast range of dynamics. For an animal, all of this combined results in vast scarcity of opportunities to sit around and do nothing. Which is probably why brains appeared on the scene.

One way a species may deal with variety is via close adaptation to a niche environment that doesn't change any faster than genetic mechanisms can re-adapt the genetic program. This way a species survives by having sufficiently many offspring such that the changes in DNA keep up with the changes in the environment. If the changes in the environment aren't too abrupt, variations in the offspring enable a sufficient number of them to survive and have offspring, keeping up an environment-driven evolution of the species. This is the main way for a vast majority of species on the planet to survive. This is the mechanism that Darwin observed in the Galapágos islands and called *survival of the fittest*.

There is another way to deal with variety, one that is a bit kinder to the individuals of a species. It requires the species to develop mechanisms that enable its individuals to do it without relying on DNA modification—what we call *intelligence*. Intelligence allows the individuals of a species to adapt to major changes in their environment after they are born, by changing their behavior in response to a changing environment. This has the advantage over other methods for ensuring survival, such as the shark's sharp teeth or the cockroaches' hard shell, that it allows addressing a wide range of phenomena on timescales

faster than genetic programming allows. When a situation has been encountered once, a memory of the encounter is formed; if the scenario turned out badly the situation is subsequently likely to be avoided, if it turned out well similar responses are likely to be used next time in similar situations. Flexible adaptation of this kind is often referred to as "learning"—although as we shall see, this term names not a single one but a set of processes, each not nearly as simple as they may at first seem. With sufficient amounts of intelligence, or should we say the right *kind* of intelligence, groups of individuals can even create new things that never existed before – such as the book that you are now holding – to address new needs, threats, and whimsical wishes.

In a world where things change at various rates, some periodically, rhythmically, as in the tides, some seemingly randomly, some semi-randomly, and some through long-term non-random causal connections hidden to the naked eye, even a tiny amount of memory and intelligence will increase the likelihood of spawning offspring. In a world where intelligent or proto-intelligent species fight, small advantages in intelligence may enable slight increases in breeding rates, which over millennia amplifies such traits in the population. The genes of those who solve survival-dependent questions better and faster in harsh environments get selected; creativity and intelligence go hand-in-hand in addressing life-endangering threats, taking advantage of unanticipated opportunities, and identifying how old skills can apply in new ways to new circumstances.

The world doesn't ever repeat itself. Even if everything you consider relevant were to be copied, down to the smallest relevant minute detail, from yesterday to today, the numbers of your calendar would still have changed, arms of the clock will be moving—a *new* day is a *different* day. So all intelligent minds, to deserve that label, must come equipped with powerful methods for managing variation—and managing the variations of variations: When is difference too big, too small, for which purposes, on which dimensions? A sound similar enough to a police siren causes a driver to slow down, or not, and possibly participates in the death of a passenger as a result; a voice different enough from her father's may result a child to check that it's really him or else risk walking away with a stranger; noticing a pattern different enough from your knitted scarf in the wardrobe at the restaurant prompts you to return it and look for yours, or else walk away the accidental thief. Examples of highly complex data at the center of our everyday cognition, these are nevertheless among the simplest examples one can find for the things a human mind does hundreds of times per hour of our waking life. If we couldn't do this reliably our lives would, in all likelihood, be significantly shorter.

Intelligence is thus a mechanism to handle the enormous and perpetual variety on Earth, and is therefore a major driver of the emergence of intelligence. It is also a major reason for why anyone would *want* to create an intelligent machine.[1]

[1] It is not my intention in this book to explain here why or how intelligence evolved – this is a more fundamental question involving a deeply complex interacting set of processes that biologists, physicists, computer scientists and philosophers are still grappling with. Here it suffices to illustrate how certain conditions on Earth made task-environments where intelligence was beneficial.

1.1 Ours is a Kind of World

Worlds in which the constraints of physics are different than in the world we live in is of course possible to imagine—as authors of fiction in fact do consistently. Fantasy may be of interest at times, and even useful for helping us come up with new ideas and exploring new viewpoints. However, even though human intelligence is capable of flights of fancy, that same intelligence is subject to the constraints prescribed by the laws of physics. And since we are in pursuit of a scientific theory of mind here, our concern here is precisely this: *actual intelligence* in an *actual world,* one that is subject to the way time, matter, and energy in the physical world behave.

For an omniscient observer with perfect knowledge of Earth's physical laws and variety, one that could see all of Earth's variables at once, at every instance of every second of every day, all future states of the world would logically follow directly from the prior state; and nothing would seem unpredictable or new. For such observation to be possible, however, the observer would have to permeate the world with sensors—something with which to record these states at every point and every instance at every level of detail. This is of course a practical impossibility: To start, the amount of storage required for all the details would require as much matter as the sum total of the matter making up the world being observed—and certainly more if the world's transformation history was to be stored. This is of course *not* the solution that nature found for preserving the life of individuals of a species. For that it came up with several tricks, most famously the one we just mentioned above, intelligence.

As comedian Steven Wright famously said, "You can't have everything. Where would you put it?" Curiously, it is the very *im*possibility of omniscience – and its cousin omni*potence*, an equally impossible proposition – that in part makes *intelligence relevant* for life on this planet: If we cannot sense everything at once at all times we need to *select* what to sense at any particular time. A memory can alleviate some of the limitation by storing that which was sensed *before now*; with memory comes the need to decide what to write and read from it, and when. This makes up part of what we colloquially refer to as *attention*.

The physical world contains organization at many scales of time and space – we say it is semi-organized at many *levels* – that present itself to us as patterns we can measure, whether with our eyes, ears, or special measuring devices. Clusters of things – objects – can be decomposed into smaller ones, to parts, materials, atoms, and quarks, creating a *hierarchy*. Both the organization and hierarchy are semi-structured; the role of science is to make it increasingly structured by explaining how certain things relate to others, in an expanding quest to ultimately explain anything and everything.

We can think of this organization as a set of hierarchies: The vast majority of variables we can measure, including those we can identify with our senses, depend on many other variables, at different spatial and temporal distances. Pick any variable of interest, e.g. the wetness of the lawn, and ask yourself what other variables this fact depends on. We can start with the enormous number of droplets reflecting light to allow you to see them as such. The light, in turn, depends on a vast number of processes. Lastly, something must have made the lawn wet—which alone will include dozens of complex variables. Further, as you walk onto the lawn to check how wet it actually is, your vantage point of the lawn, and

the millions of droplets, is changed, yet you still see it as the "wetness of the lawn". The complex changes in the lawn's shape as you move about in three dimensions, the differing projection of it on each of your eyes, etc., are ultimately a matrix of a huge number of variables that have particular relationships to each other. Such dependencies are called hyperstructures and they produce what statisticians refer to as covariation between variables. These covariations are what our minds make use of to create knowledge of the world, so that we may move about and achieve goals.

Hyperstructures of interest to humans include the production of rain from clouds, the growth of grass, trees and fruit, the flow of water in rivers and kitchen sinks, the movement of people in buildings and the jungle, the driving of cars in streets and on pavements. As you hold your cup of coffee and bring it to your mouth its projection on your retina changes constantly; our stereo vision allows us to extract a hyperstructure that predicts its shape so that it doesn't surprise us. Even though we may never see the cup exactly the same way twice (you're always a bit further away, a bit further to the left, the lighting slightly different—there really is virtually no chance that you will match your relative position to the cup to the level of your visual acuity on two different occasions) the cup just appears to you as a cup.

Hyperstructures also allow us to recognize what we typically think of as 'context' or 'environment' because any measurement can enter into a covariation matrix—and if it helps us distinguish between situations where some action (say taking the subway) may be performed with some success (say before 1 am and after 5 am) then we allow it to enter. It may not be technically part of the task, but it is part of what makes it possible for us to do the task successfully.

In addition to hypestructures helping our minds make sense of billions of variables that never repeat exactly the same way twice, the mind has mechanisms to assess similarity. With piecemeal, incremental experience we are thus able to determine how similar something must be to something else for us to predict its usefulness, danger, or irrelevance to ourselves at each moment. Similarity is important for worlds, because a completely random world is completely unpredictable—nothing would ever be related to anything else, by definition. In such a world intelligence is pointless: A mind could not extract any regularity and hence it cannot learn. Even if it knew something it could not know anything about the random world that would help it in any way—the only thing it could know that would be true is that in this world it cannot know anything. So this knowledge would leave it in the same place as if it knew nothing: It could not do anything. Clearly, the physical world is not completely random.

Similarly, at the other end of that spectrum, the world does not appear to us as completely deterministic either. I say "does not appear to us" because what matters here is that *to us* the world is not completely predictable. To explicate this distinction between a world that *is* non-deterministic and one that *appears* non-deterministic we can do a short thought-experiment. Imagine the smallest entity that we will discover in the future. Let's say that will be called "floob". A floob has two states – it can be "on" and it can be "off" – and when measured it is always found in one of those two states. However, no variable has been found that can predict its state, and its state has not been found to correlate with anything

that we have ever measured. Scientists now seemingly have two choices: They can say that this is a truly random variable, and that the world is essentially random at its core. Or they can say that so far, they have not found the variable(s) that may predict its outcome, but that they'll keep looking. And the fight about whether the universe is truly random or not goes on, until such a variable is observed. And that seems to be the end of it. Except that they simultaneously then discover another-even-smaller entity whose state seems to correlate – again – with nothing. Both claims turned out to be true. So, is there any hope for the question of whether the universe is truly deterministic or inherently random?

To get to the bottom of this we must talk about *substrate*. The substrate of something is that which it is made of, and thus depends on. The substrate of matter that we understand reasonably well is chemical bonds—ways of binding atoms together; the substrate of computations in a desktop computer is transistors and electricity—the stuff that runs the computations; the substrate of human thought – the stuff that produces and sustains cognitive processes – consists of electricity, chemistry and neural communication (and probably other things). When talking about a phenomenon in the world at some level of abstraction, like computations in computers, we can relatively easily identify its most relevant substrate (electric charges in electronic circuits). Not so with the substrate of the universe, because no matter how far down we dig, the smallest entity or phenomenon we find at any time *might* not be the last—and whether it is there is no way for us to know—ever.[2] Let's imagine that the smallest thing ever observed, the hypothetical "floob" mentioned before, will actually turn out to be real and be discovered sometime in the near future. Then nothing smaller is found for a million years. Does that mean it really truly is the smallest thing in the universe? No. That does not mean that smaller things don't exist: There is no way to rule out that phenomena smaller than the floob exist that are simply still *unobservable*; whether they are unobservable for microseconds or millennia doesn't have any impact on whether they exist or not. It may make it less likely, but it cannot rule it out for sure. But the problems is even deeper: Even if the floob actually and *truly were* the smallest entity that our universe is made of, there is no way for us to know. We cannot penetrate the substrate of our own existence. Why?

What do we mean by "substrate of our own existence"? Outside our universe there may in fact be some other machinery – fully deterministic mechanisms, or not, who knows – that remains forever out of our reach, because it is part of the very substrate of the universe, and thus the substrate of *our own existence:* Just like a simulated character in a computer game cannot ever penetrate the simulation to take a look at the substrate – transistors – responsible for *its* existence, we cannot ever go outside the "program" that is our universe to see what it "actually runs on."[3] This is why we cannot ever be sure we have found the

[2] It gets even stranger still—see Bell's Theorem, https://en.wikipedia.org/wiki/Bell%27s_theorem — accessed July 3rd 2019.

[3] Computer scientist Ed Fredkin addressed this in his 1992 paper *Finite Nature*, where he explains how a simulation of physical airflow relies on a computer that is running the simulation; the computer is not in the simulation, it is outside of it. Therefore, the simulated airflow cannot affect the computer, it exists because of the computer. Thus, if we implemented a measuring device in the simulation that could measure anything it would still be limited to data in the simulation, not outside of it; it could not measure anything the computer uses to run the simulation of the measuring device—bits and bytes, memory locations, instructions, etc. that the computer uses would be out of reach of the measurement device. By the same token, any entity in the physical universe – be it airflow, humans, or some other physical thing – cannot affect or in any way measure

"bottom turtle."[4] It is also why speculations on whether the universe is "a simulation" will forever remain a philosophical or religious question at best, and a completely meaningless one at worst.

It is also the reason for why all knowledge is non-axiomatic. Unlike mathematics, where some statements are given and certain – its axioms – the universe does not offer us anything of the sort. The closest we have come to axioms are probably Einstein's Theory of Relativity and Descartes' statement "I think, therefore I am." Descartes' is more fundamental because it cuts at the very core of our being: It says that the most fundamental knowledge is the knowledge that we have knowledge. He could also have said "I experience, therefore I am," because what we have, at the end of the day, when we *think* is the *experience of thinking*, and traces of *memories of having thought*. However we formulate his idea, its basis is indisputable; this is the most fundamental and certain unit of evidence we have about the world—everything else is inferred, e.g. that other people have minds (an inference based on assumption of similarity), that there is some "stuff out there" that does not cease to exist when we stop thinking about it, that the brain is what causes us have these thoughts, and that our scientific theories "really describe" what is "out there," and so on. Looked at close-up it seems, admittedly, rather a feeble a foundation to base the decisions of a lifetime on. However, it is what we have to go on, and we might as well make the best of it.

Making "the best of it" means, primarily, trying as best we can to make sense of it, e.g. explain it via a comprehensive and compact set of models that dictate how it works. On that note, getting back to the question of determinism, we can conclude from the preceding discussion that the world will probably always have, from our vantage point, an element of randomness, because there will always be processes and variables that we cannot access—factors that are outside of our reach because they are the factors that underpin our very existence.

In summary, Earth is a kind of world with various kinds of features, each of which make it a specific kind of world. The features of the world we have covered so far include:

- Semi-structured. The world is poised between complete randomness and complete determinism.
- It is hierarchical in that sets of a multitude of measurable variables correlate with, and are in many cases completely dictated by, higher-level variables.
- We can reduce the randomness by peering into the inner workings of the world and make models of it that explain and predict.

1.2 The Arrow of Time

People have a difficult time with time. Time itself is so ... so fleeting, and temporal.
One minute it's there, the next it's just ... it's just not.

—Reggie Watts

what "runs" our universe. Therefore, there must exist a boundary beyond which we cannot reach when trying to find out out what the universe is made of.

[4] …

Flowing rivers, the falling of snow, the movement of the wind over time—all point to a defining feature of the world all intelligent creatures live in: *Time has a direction*. Luckily, the flow of time, and the interactions between phenomena in the world, are subject to *rules*. These rules rule our world – literally – by determining what may and what may not happen. Without this single fact intelligence would not only be impossible, it would not even make sense.

The fleeting of time; the cold snow; my increasing hunger at the clock ticks towards noon;—these are very practical problems. Indeed, it should not come as a surprise that intelligence is a very *practical* phenomenon: Due to the enormous variety of stuff on Earth, and an ingrained drive to perpetuate a species, intelligence is a solution that allows individuals of a species to survive in the face of practical everyday challenges. To cut a long and complicated story short, *intelligence is a practical solution to practical problems*.

Intelligence is rooted in a constraint – a negative goal – to avoid extinction,[5] in a complex, detailed and non-repeating but somewhat self-similar environment. It couldn't be any other way—otherwise we would simply not be here. Yet this is no simple feat to accomplish. Would a mind not require a whole tool chest – nei, a warehouse – of specific solutions to specific problems to succeed at such a complex task? The "intelligence solution" is different: If all we had in our head was a warehouse for experiences, Earth's diversity would quickly fill it up—no matter how big the head. Instead, intelligence works differently: By creating models of the general features of the world, and allowing the on-demand construction of solutions. To be sure, a lot is remembered and stored (in "our mental warehouse", if you will), but it is not monolithic specific solutions to specific tasks or environments; instead, it leverages the regularities in the individual's experience of the world to compress it. Instead of storing everything (and thus taking forever to retrieve it when needed), intelligence provides mechanisms to generate solutions *on the fly*.

Granted, intelligence allows an agent to do a lot of other things besides avoiding extinction. This is especially true for the human species, for which we can draw up an enormously long list of weird and wonderful "extracurricular activities" that people spend their time on – things that one may be hard-pressed to categorize as sub-goals of survival but we pursue nevertheless – whether it is because we find them interesting, fun, important, or see them somehow as "necessary."[6] Because general intelligence is in fact general, it should not come as a surprise that it can be used for many things that are only marginally related – and even not related at all – to our species' survival. Be that as it may, a large part of what we do, even daily, is to solve problems, whether it is problems posed by ourselves to ourselves – like what-to-have-for-lunch-if-not-that-convenient-sandwich-shop-down-the-street-that-we're-sick-of – posed by the environment (an oncoming car, finding a new job), or by others (a teacher demanding your essay by Monday).

Any agent living the physical world must deal with the flow of time. The here-and-now is a fleeting instant; no matter how quickly you utter the word "now" the statement becomes a

[5] Stephen Pinker, *How The Mind Works* (1997).

[6] Pinker went so far as to say that "…the apparent evolutionary uselessness of human intelligence is a central problem of psychology, biology and the scientific world view" (1997, p.300).

thing of the past in an instant. To learn from experience requires, therefore, a memory of the past. The smallest temporal instant that a human can sense – the human temporal acuity – is around 10 milliseconds (you can try this by playing two "tick" sounds closer and closer together in time—as they approach 10 msecs they blur into a single inseparable "tick").[7] Part of this is due, of course, to the time it takes to measure the energy coming into our sensors. This means that for any perceiving agent there is an interval that we call "now" in which nothing is known about the future; as it rolls in it gets measured by our perceptors, turned into neural signals, and – at some point later – conscious experience and information that we can act on. For humans the "now" is no shorter than 10 msecs, in many cases longer. The further into the future we think the more uncertain things become; the further back in time we look the more uncertain it is how things actually came to be.

[7] Hearing has the best temporal resolution of our senses; vision tends to blur things at 20-30 msecs, touch even larger.