

The Rocky Road Towards Defining the Mind

Sandro Skansi

Abstract

This study scrutinizes the essence of intelligence through the lens of search theory, enriched by philosophical insights and computational paradigms. We critically analyze Herbert Simon's foundational idea of intelligence as search, revealing its limitations in capturing the complexity of human cognition. Emphasizing the role of imagination – a neglected aspect – we explore how it simplifies intricate realities by reshaping search spaces through conceptual frameworks and classifications. Our exploration navigates between materialistic reductionism and dualist views of the mind, scrutinizing neural mechanisms versus the intuitive aspects of mental phenomena. Ultimately, we advocate for an integrated perspective of intelligence that goes beyond algorithmic problem-solving to embrace creativity and the nuanced depths of human thought.

Key Words: mind, dualism, mental phenomena, computation, cybernetics, logic

228

DOI: 10.5281/zenodo.14272629

Introduction

Since the dawn of civilization, the human mind has puzzled philosophers. In a world filled with mysteries and magic, the human mind seemed more accessible to research than nature, but today we feel quite different. We are way more sure about our competences and quite fond of the so-called “scientific method”, which has the role of fine plates and cups always on display in a glass cupboard but rarely used. Although we know much less about what we consider to be the world than the ancients did about their world, the study of the mind has proven to be particularly difficult. Since at least Plato (Plato, 1966), it seemed that the human mind was something substantially different than, say, a man's finger. It seems as if the two have almost nothing in common. The idea was solidified in the Christian idea of a

¹ **Corresponding author:** Sandro Skansi

Address: Faculty of Croatian Studies, University of Zagreb, Borongajska 83d, Zagreb, Croatia

e-mail ✉ sskansi@fhs.hr

soul and finally with Rene Descartes dualism (Descartes, 1998): there is the mind, and there is the body.

On fingers and toothaches

That will be enough for an introduction – let us dive into the peculiarities a bit. Consider a human finger, which blindly obeys our every wish, does nothing on its own and can give us a sensory input from the environment. As part of the sensory input, it delivers pain as well, and for most of us, that pain cannot be avoided by "refusing a delivery". If a painful event happens to the finger, the mind will know about it. It will never be able to ignore it.

The mind, in contrast to the finger, seems to be in control. It is the mind who sends out instructions, and it is the mind which understands sensory input. But when we dig just a bit deeper, the human mind looks rather different. Read the following sentence carefully, and do your best to obey it:

Do not think of you favorite food

This is a paraphrase of the classical argument of George Berkeley, called the master argument (Berkeley 1957: §22 and §23). People have a hard time not thinking about something that is presented, and this property can be called "spontaneity" – the mind does what it does all on its own and does not stop. The finger has no such problems, "DO NOT TOUCH THE HOT WOK" is very clear and not at any moment does the finger come close. If it does, it does because the mind said to do so. Yet the mind, who has seemingly perfect control of the finger, seems to have little control over itself.

Our mind controls the body, and it is quite good at it. It also produces reactions to the environment. Controlling these is harder. But turning itself off, or not accessing something that was delivered by the sense seems all but impossible. The mind seems filled with these mental states that are generated automatically. It has become customary over the years to call these mental states "subjective", but Kant (Kant, 1781) argued that most of them are not "subjective" in the usual sense of the word.

Imagine you are tidying your work desk. The way you place things seems completely arbitrary. Yet you would be surprised that most people tidy it in exactly the same way. Why is this so? Because, as Wittgenstein put it (Wittgenstein, 1953), we share a form of life that is very specific. A sociologist might not need a calculator, while an economist does, but they both need pens, color pens, mobile chargers, cable ties, paper hole punchers, etc. What is interesting is that they will need some of the items more often and other items less often, so some of the items will be closer, and some will be deeper. Even though

we have a free will to choose any arrangement we want, our form of life (which is actually our interaction with the environment) dictates we have only one optimal way. There might be situations where there are two or more equally good ways, meaning that it is hard or impossible to determine which of them is better, but if you can compare any two arrangements to see which is better, then there will be only one optimal arrangement dictated by their work, i.e. the form of life they live through.²

It seems that our mind has a way of escaping “subjectivism”. Even though it is not at first glance clear how exactly, it seems that the environment has a lot to do with it. But how did the environment shape the human mind? As with most things, evolution seems to hold the key. The basic interaction is via pain. An animal avoids things that bring pain, as pain was associated with danger. But as anyone who went to the dentist knows, the world is much more complex than that. There are cases where our mind has to embrace pain. It might be tempting to say that the mind, the interpreter of the pain signal from the finger (or tooth), simply reinterprets the pain away, but this is not what happens. It seems that some (higher) faculty says to the mind not to ignore, but to endure the pain. It might even make the fear of pain worse by knowing what follows, and yet the mind stands its ground and endures the pain.

While the brain today is no longer the mystery it was even a hundred years ago, we still have a problem describing the mind. There is an interesting and comical idea on how to do it, reminiscent of Gilbert Ryle's famous argument (Ryle, 2000): simply imagine a tiny human living in a person's head. Joking aside, we could probably modify the idea to make it a lot more interesting: imagine a whole society living in a person's head. This was the idea set forth by Marvin Minsky (Minsky, 1988). Minsky proposed a model of the mind which uses mindless agents, each performing one task (or part of a task) and their synergy constitutes the mind. It is easy to miss Minsky's dualism put forth here: if the agents themselves are not the mind, but their interactions form the mind, then the mind is the very hierarchy that constitutes the society, along with the processes being implemented.

Historically speaking, Minsky's theory is new, modern. Yet it still talks about the mind in metaphors, not in neuroscientific terms. Also, Minsky's theory, along with all theories of mind, seems annoyingly descriptive: trying for the hundredth time to describe what the mind does by using extrapolations from processes we (supposedly) understand better, such as society. It is painfully obvious that these theories are at best only crude sketches of a complex reality.

²This echoes Ashby's (1956) homeostasis and organization as a part of the definition of a machine. See Greif and Šekrst (2024) for a philosophical insight into Ashby's rediscovered manuscript from 1941 on dynamic organizations and evolutionary adaptability.

So far, we have been talking about fingers and teeth on one hand and the mind on the other. Why the charade? Why not talk about the brain right away? Well, we could, but the brain, viewed from a dualist standpoint, is just another finger, but the implications when talking about the brain seem to cloud the discussion. Most people adopt dualism because it allows them to use a specialized, yet simple language to describe the mind and its processes, and not because they are inherently dualist, which becomes most apparent when the brain enters the discussion – for a true dualist, the brain has more in common with the finger than with the mind.

The language which the dualist uses includes the words “mind”, “feeling”, “soul”, “intuition”, denoting elements of the mind. This is in stark contrast with the non-dualist would need to define them in terms of brain-elements. Yet, for all the linguistic assets available, most people are not dualists in the real sense. Even religious people, who should be at the forefront of the dualist effort, often believe that all elements of the mind will at one point be explained and localized in the brain. This in turn generates strong expectations towards cognitive science and neuroscience who are tasked with finding the “brain centers” of all elements of the mind. This ongoing dialogue reflects broader societal expectations that cognitive science and neuroscience will eventually pinpoint the specific neural correlates responsible for each facet of human experience. Amidst these debates, the exploration of reasoning emerges as pivotal.

Reasoning and animals

One of the most fundamental things our mind does is reasoning. It is worthwhile to explore how we came to be the “rational animal” of Aristotle (Aristotle, 2012). Descartes (Descartes, 1641) noted that even though the original definition was not so simplistic, in its received form it begs the question of what do we mean by “animal”. This seems rather simple, but in fact it hides a sea of complications. The very concept of “animal” is not as well-defined as one would hope.

With Aristotelian definition it is implied that humans are animals, which is in itself not problematic as some would like us to believe. Actually, everything else is the problematic part. What about extraterrestrials? Would they be considered animals? The fine detail here is that suddenly we are faced with redefining life as we know it? Should the hypothetical aliens turn out to be silicon-based (Pace, 2001) instead of carbon-based, we might need to reconsider our very own computers to be “animals”. Notice that the fact aliens are not proven to exist is irrelevant. As long as there is a possibility of such aliens, our definitions are flimsy, i.e. their wellfoundedness depends only on the contingent fact that we have not yet discovered aliens.³

³As a weird side note, there was a philosopher who took the rational animal definition and

Returning to the definition of humans, the question of humanity via rationality seemed perfectly fine until the advent of computers. It is a bit unclear when this happened in this context, but today we have computers which can do most higher cognitive processes. In fact, we could pose a conjecture:

There is nothing cognitive which is uniquely human. Every cognitive aspect of the human mind is either present in the animal mind or in computers.

The keen reader might see that the actual Aristotelian definition might still be intact, due to the “animal” in the *genus proximus*. As noted earlier, the problem with all the approaches is in which we cling on to the *genus proximus* is that they seem accidental. Along with “rational animal”, Aristotle used a second, comical, definition “featherless bipods”. Thinking alongside Quine (1980), one cannot help think that one definition is serious and substantial, while the other is comical and accidental, but it seems that with the advent of computers, and basic AI, even the former definition is by and large as accidental as the latter, albeit slightly less comical.

But let us assume for a while that the *genus proximus* is not problematic. In that case, when considering humans as rational animals, it seems that our intention is to define humanity as rationality. And this way to capture humanity has gone unchallenged for centuries. Throughout history it was sufficient to say that humans are those entities who are capable of learning to, say, multiply numbers. Learning to multiply is not something as clear as one would like to think, and Saul Kripke (1984) has made a household name for himself (in part) for analyzing this phenomenon.

Kripke noticed that there is something quite odd in rules. Take addition for example, which can be thought of as a set of rules, which apply to all natural numbers. A little bit more precisely, addition is an algorithm (which is a set of rules), which takes as an input two numbers and outputs a third. But, as most people know, there are infinitely many numbers, and you could not have been instructed in adding them all. You were instructed with a couple of examples and then you were expected with some hand waving to extrapolate the rules. Imagine that by some accident you were taught numbers up to 999 and all additions you have seen performed never used anything above 999. How would you know that $1000+1234=2234$? You could do a digit by digit addition. But what about $9000+2222$? It is hard to

modified it to make it absurd: Ernst Cassirer (1944). In his essay, he wrote that humans are “symbolic animals”. This definition is a prime example of philosophy gone wrong. How can we have a definition of “symbol” that excludes animals we know? Bees use flight patterns, ants communicate with pheromone patterns, birds have courting dances, and greater apes know how to compose words with letters. Symbols are found everywhere in the animal world, and they cannot constitute the differentia specifica for humans, unless one wants to consider chimps, bees and magpies as “human”

imagine this now that you know how to add, but here is something that might help. Think of the words “ten”, “hundred”, “thousand”. How would you, from that, infer the name of 10000, 100000 or 1000000? Why would it be “ten thousand”? Or even more intuitively, why would it be “million” and not “thousand thousand”? If you have never heard it, you cannot know it. So if you do not even know how the result of $9000+2222$ is called, how do you presume you know what it is ontologically? How do you know it is 11222? How do you know there is not an exception in the rule expansion, similar to how we do not have “thousand thousand” but “million” in the rule governing the naming of numbers?

These ideas bring us to our first paramount topic, *free will*. If we had a system which is governed by two rules, and operates on say 20 different input entities, it is much easier to list the rules and entities. The process itself is naturally deterministic. Of course, as the complexity of the system grows, it might be helpful to describe it stochastically, even though we know that the underlying reality is deterministic. A simple example is a coin toss. In theory you could write out very complex equations which tell you exactly what will be the outcome of each individual toss, but it is far more simple to describe it as a probability distribution. What if we increased complexity even further? What if we have a small number of elements, and a huge complexity of rules? Then it might be tempting to describe this system as “free”, not in the sense that it is indeterministic, but in the sense that is simpler to describe the system as “free” by saying what it did than by saying how it did it. Such freedom can in fact be reconciled with determinism by claiming that the “freedom” is actually the result of the huge complexity we have when we try to approximate, let alone describe, the mind as a set of rules. This might be seen as a plausible route to compatibilism. The indeterministic and incompatibilist version of free will does not merit more than this sentence.

The mind is obviously not “free” to ignore certain stimuli. There is a huge tradition which claims the mind as free in terms of choices, and we have argued that this notion is increasingly nuanced in light of advancements in artificial intelligence and our evolving understanding of cognitive processes. In conclusion, while advancements in AI and neuroscience provide insights into the mechanisms underlying human cognition, they also prompt a reevaluation of our concepts of rationality and free will. The exploration of these topics continues to evolve, challenging us to reconsider what it means to be a “rational animal” in an increasingly complex and interconnected world shaped by both biological and artificial intelligences.

A first attempt at building an artificial mind

To the modern mind it seems that not many things constituting the original, ancient, non-neurological, philosophy of mind still hold. However, the initial underlying proposition that the mind is complex and linked to the brain holds more firmly than ever. In that vein, the attempt of Pitts and McCulloch⁴ from 1943 seems a noteworthy attempt to establish a new philosophy of mind. Pitts and McCulloch noticed that, up until that point, the major problem in the mind-body investigations was Cartesian dualism. Dualism was all too keen to serve the soul-finger analogy. They took the simplest path, the finger, whose function was described mechanically, and the soul, which was described almost theologically. By doing so, the Cartesians made the dualist distinction seem more fundamental and the divide greater. Not only that, but they, in a way, put forth the idea that the finger and the soul (representing the mind and body) should actually be studied by different sciences, located on different ends of the spectrum. One was almost the subject of medicine, while the other was almost theology. The languages those sciences used were also vastly different, which made any “interdisciplinary” research next to impossible.

McCulloch and Pitts (1943) wanted to bridge the divide. The first step was to take the most mechanical faculty of the mind, reasoning, and the most esoteric element of the body, the *neuron*. Just by doing this, modern philosophy of mind sprung out of the bottle. Their basic methodology was opposite to dualists. Even though they did not know at first how to connect the mind (reasoning) and the brain (neurons), they knew that the best place to build a bridge is where the two islands are closest. This by itself was enough to influence cybernetics in the late 1940s, artificial intelligence in the late 1950s and analytic philosophy of mind in the late 1960's.

The paper McCulloch and Pitts wrote in 1943 is not taken by many researchers to mark the beginning of the modern philosophy of mind. This is a major injustice, since McCulloch, among other degrees, had a degree in philosophy, while Pitts' only degree, Associate of Arts, was finished under the supervision of Rudolf Carnap (Gefer, 2015), a well known analytic philosopher at the Department of Philosophy of the University of Chicago. One could only speculate why Pitts and McCulloch were not venerated as the two fathers of the new philosophy of mind in the same way Edmund Gettier (Gettier, 1963) would be championed as the father of the new epistemology twenty years later. Perhaps people today think that it is enough that these people are venerated for being the founders of artificial neural networks and deep learning.

⁴For the original argument see Pitts & McCulloch (1943), but for a great modern reexamination, see Perkov (2020).

McCulloch and Pitts wanted to show that a faculty of the mind, reasoning, can be learned from examples by a simple set of rules which represent the functionality of the neuron. By doing so, they made possible the definition of an artificial neuron which could be implemented on a computer. It was, for all intensive purposes, a purely logical model. It could take two inputs, which were to be multiplied each by its own weight and added. If the sum reached a certain threshold, the neuron would output 1, and 0 otherwise. In this way, basic logical functions AND, OR, NOT, representing the most basic forms of reasoning which the mind does, could be learned.

This tiny model makes a huge change, for man is not a “rational” animal by metaphysical necessity: humans learn to reason, and by consequence, and Pitts and McCulloch showed how exactly this happens. These examples could be given by a teacher, but they could be also abstracted from nature. From 1943 onwards, humans are no longer theological creatures, whose essence, rationality, was endowed to them by a mystical creator, but they were animals interacting with their environment, more similar to ants or elephants, than to the gods of old.

Reasoning revisited

To see what makes reasoning so special, let us take a digression to the birth of the computer. Most people today consider Alan Turing to be the father of modern computers with his 1936 lecture (Turing, 1937). This idea is highly subjective, but in a very important way, it was not Turing but Claude Shannon who made the crucial step in his paper (Shannon, 1938). The people who see Turing as the father of modern computing often use the argument that he defined what an algorithm was. This is true, to a degree. First, Turing was not the first person to define an algorithm (as one can easily infer from the very name “algorithm”), but he was the first to define it as a mathematical entity, so precisely that negative results could be proved. Take any problem you want, e.g. the problem of sorting books by size. To do so, you will use an algorithm, and this algorithm can be just a procedure written on a post-it. But what if you have a problem for which no algorithm can exist? In this case, saying that you do not have an algorithm does not suffice. What you need is a very precise definition of the problem, and what an algorithm is and then you can hope to prove that for that problem, no algorithm can exist. Of course, a slightly modified problem can have a simple algorithm. This is exactly what Turing did. This is an important result, but it connects two ideas, the idea of the problem, and the idea of the algorithm (i.e., existence or non-existence of the algorithm).

Claude Shannon did something much more important: he showed how to connect an idea to a concrete, material, machine. In his work (Shannon, 1938), he showed how to implement logical

operations using the hardware of telephone relays. This in essence was the birth of the software-hardware distinction, and bears a strong resemblance for the modern person to the mind-body problem. In fact, we have heard on numerous occasions people explain the mind-body problem by using the software-hardware analogy. But the important part here was that he made it possible to have electronic calculating machines. There were “programmable” machines before, but this time it was very different. What made it different was the possibility of handling infinite numbers and calculations with finite switches and wires.

So how can finite computers use infinite structures? The structure we want to showcase is $(\mathbb{N}, +)$, that is natural numbers with addition. Subtraction, multiplication, division follow quickly. Order is interesting, and we will sketch how order can be defined. So first we need to take care of the basics. We assume we have a machine, capable of handling boolean operations AND, OR and NOT. Boolean values are just 0 and 1, and results are also just one value. For AND and OR it is simply a serial or parallel electronic circuit. NOT is a bit more tricky, but it can be made to work with a little ingenuity, by declaring 0 not to be the absence of electricity, but e.g. low current since the problem with NOT occurs when we want to (trivially) represent NOT with a switch. Turning the switch off does change 1 to 0, but turning it on will not convert 0 to 1. All this can be solved with a simple power converter instead of a switch.

Once we have a machine doing Boolean logic, we can easily extend it to do addition on binary numbers. First we need to create binary numbers, and this can be done by simply making all strings of 0 and 1 up to a certain length and simply removing the ones starting with a string of 0. Note that it is enough to remove just those whose first digit is 0, since it will automatically remove all strings which are not binary numbers. Also note that by ordering them by length, we will get some order: for each two numbers a and b , where a is a shorter string than b and a power of 2 named c that sits between them, we have in actual order $a < c < b$. But we can do a lot better than this by simply seeing that order on boolean values means that $a < b$ is true if and only if $a=0$ and $b=1$, and one can quickly see that this is simply defined by a formula $not-a$ and b .

It takes a bit of care to extend this to binary numbers, but it is an easy exercise for most readers. To define binary addition with boolean values we will need the function AND for two and three arguments, the function XOR and the function MOST (all of them capable of taking three arguments). The function $MOST(x,y,z)$ is simply xy or yz or xz . We will also be needing the idea of a carry number which stores values to carry on. By applying AND to the last two digits of the numbers we want to add we get the last digit of the carry number. By applying MOST to the last digit of the carry number, the second to last digit of the first number, and the second to last digit

of the second number, we obtain the second to last digit of the carry number, and so on. The same procedure, but using XOR gives us the result. The whole procedure terminates rather quickly, since we just go from the back to the front. Even a number of length 100 needs only around 200 steps to compute, and note that these are bitwise binary steps, which are very fast.

Glossing over the details, this is how Shannon showed that it is possible to have an electric machine, implementing (finite) Boolean algebra, to do calculations with infinitely many natural numbers. It is impossible to overstate the metaphysical importance of this result: a finite machine was able to exploit peculiarities of Boolean logic, and this in turn opened a way to explore infinite mathematical realities, and thereby reconciling the material world of physics and the ideal world of mathematics.

Reasoning about reasoning

Notice how in every one of the cases discussed so far, logic plays a special role. One might be tempted to say it is some esoteric *ordo rerum*, but the truth of the matter is that logic, similarly to the Golden Ratio, forms the most basic structures, which both the material and nonmaterial world share. This makes logic the central place to study the metaphysical interconnections of reality.

Even though logic seems to be the most basic abstract structure in nature and the world of ideas, it seems that a third world, separate from the first and second, should be considered: the mind. Even though it would eventually coincide with one of the two (or both), there is something special here. This can be most easily seen by falling back to the McCulloch and Pitts (1943) paper. The natural world here would be implementing logic as a computer, the “world of ideas” would be implementing logic to define the (artificial) neuron. Finally the mind would be the one to learn logical functions from examples. In this tripartite world, the mind handles sensory data (examples) and learns (the weights).

Nature itself clearly realizes the brain in physical terms, while the world of ideas is responsible for the conceptual structure of the brain, made so that it will be able to learn. How this conceptual structure came to be is a matter of taste, some like to see it as natural selection, some have the need to pin it to a deity, but it actually does not matter. The mind itself has nothing mystical or theological to it – it is concerned with doing its best to learn from the environment, so that it may adapt to it as quickly as possible (and gain an advantage). The natural world is also not problematic, being wholly material and easily traceable to simpler organisms. Note that this trifold view of reality is not real – we are not proposing “trialism” or anything like that – it is just a conceptual tool to show a way we could look at it. Most serious research today points out that there is a single world

where everything happens, and it is just our current inability to explain it in terms of a single reality, with a single conceptual framework, and an unified language to handle it.

One could argue that logic is a special thing indeed. It exists as an academic discipline, connecting philosophy, mathematics and computer science, but also as a basic human faculty. Whether logic in this latter meaning is a human virtue (not possessed by everyone, but only by a few worthy souls) or a human sense (similar to say smell, possessed by almost everyone) is an active discussion. Even though logic as a science seems to attract very few researchers, and even fewer are able to produce lasting contributions, there is a strong intuition that logic as a basic human faculty could be considered a “sixth” sense – once the logical eye sees something, it cannot be unseen.

The idea that logic is something more similar to the five senses than to a classical mental construct is not new. Even though the idea itself might be older, it was first investigated by Kant (Kant, 1781). Most experts on Kant would probably disagree with us due to a strong tradition of separating senses and what is traditionally considered purely mental matters, categories. Kant also honors this distinction, but he does in fact provide the same reasoning for transcendentalism for both forms of senses (time, space) and logic (which he called “transcendental categories”). For Kant both participate not as a content of experience, but they form any possible experience by shaping the experience. In Kant's words, we can imagine empty space, but we cannot imagine an experience that has no spatial aspect. Even though we will propose a different theory, Kant was undoubtedly the first philosopher who saw a sameness in senses and logic, by seeing that they are both necessary for any experience to form and exist.

Kant was an optimist in thinking that logic was deep within our minds and that all of us project it into the world. For if it were so, all humans would have it, and as we can see almost everywhere, people do in fact make illogical inferences. One of the most common conclusions which is not logically valid, yet quite common among people is *abduction*:

“All cats are mammals.”

“Felix is a mammal.”

Therefore

“Felix is a cat.”

If Kant's argument, or any extension thereof were valid, there would be no explaining people erring in a common way, and formulating “pseudo-valid” arguments, i.e., arguments that are in fact not logically valid, but a larger number of people accepts them as valid.

It seems that logic would have to be either found as an additional sense (which can be refined) or in nature. In a sense, the former option also boils down to the latter: if logical reasoning is a sense which can be refined (similar to say the sense of touch), it can be refined only in interaction with something from outside, i.e. reality. This reality can in fact be abstract, but it still forms an experimental sandbox for practicing the art and craft of logic.

Suppose we want to teach someone the Pythagorean theorem. One way of showing would be to let the person measure the thingies involved. By measuring various examples one does not prove it, but gains enough confidence in the veracity of the initial “Pythagorean claim” (which cannot yet be called a theorem without a proof). Once he has measured enough triangles, he will see that no matter how many he measures, there are infinitely many more. A “proof by measurement” would have to encompass all of them. A simple, but actual proof which works for all triangles is illustrated in the image #, and it stands to reason that our budding mathematician might find it himself.

Similarly, a person might see the validity of deduction vs abduction by “drawing” a diagram as illustrated in image #. Similar to the Pythagorean theorem, once this abstract structure is uncovered, all of its concrete subcases become natural. But it has to be uncovered, and it is learned from examples and abstracted to by the mind becoming bored with the same thing over and over again. There seems to be a spontaneity in the mind, as Kant argued, which seems to warrant some investigation. The mind seems to have an appetite for causality, and this is evident not just in our best efforts to explain the world, but also in our blunders. Even though most people imagine the idiot to be the person who is clueless in terms of causality, this seems to be nothing more than a strawman. Indeed, more often than not, an idiot can be characterized by finding causality where there is none. He is all too pleased to bundle simple correlations in a causal web of his own weaving. Although the idiot is a walking caricature, every man has a bit of an idiot in him. For it is the regular man who has a spontaneous mind, which is far more hesitant, but nevertheless equally predisposed to connect the dots. Searching for patterns is what the mind does. There has been a long tradition in formal logic of misrepresenting causality, the most obvious one being equating causality with implication. The classical “If A, then B”, is a proposition which supposedly captures causality, but upon closer scrutiny it actually captures less of the phenomenon than a simple conjunction. To see this, just write out a truth table and ask yourself in what world would we accept “A causes B” to be true just because A is false?

To capture the elusive “A causes B”, formal logic will never be enough, and this means that in turn, there is no inferring to the causal structure from static logical propositions. Causality must come from outside the mind, at least in part. And that part that needs to come

from the world is the notion of exogenous time. By a slight modification of Hume's (1739) argument on causality, "A causes B" is true if and only if A is true, and *after* that is ascertained, B is ascertained to be true. The concept of exogenous time, unlike Kant's endogenous notion of time, is an absolute necessity for not only learning causality, but even for representing it. The human mind seems to require external time to correlate the internal representations of events, almost the direct opposite of what Kant claimed. This is because the internalization of propositions is a simple task, and indeed we do not reason with objects or events, but with statements regarding objects and events. Should it not be so, we would have to have two faculties of reasoning, one reserved for objects and events, and the other reserved for more ethereal matters, like concepts. Time itself comes from outside, and it is absorbed to structure propositions in a (tentative) causal network. With a tentative causal proposition, evaluated by the exogenous time, we are ready to confirm it with an experiment, where we are free to rearrange the components to make a different order, and see whether they still hold. The objectivity comes from the unidirectionality of exogenous time: if A causes B, we can repeat it many times, and even try to place B in the flow of time hoping to get A. Or try with some C and D. But the mechanism which connects them and enables causality, is the objective, exogenous time itself, as if they are objects being dropped in a river and then observing what happens. In a sense, by doing so we learn much more about the river than about the different objects we are dropping, and by building a causal network we learn objective time, not causality. Causality is nothing more than the spontaneity of our mind exploring objective, exogenous time. By being animals who seek causality, we explore exogenous time, and handling representations of objects and events in a causal net we are doing the same as we do when programming with functions and objects: explore the flow of objects in exogenous time.

It can be stipulated that our brains are hard-wired to find causality in the world. This is an interesting perspective which avoids some problems. By having this Chomskyan view on causality,⁵ we create a specific kind of dualism: there is the world with its material complexities and its causal relations, and then there is the brain, with its preconception of causality and an obsession to find it. These two causalities evolved to be in sync, but they are different. The one in nature was governed by laws of physics, while the one in the mind was a mere accident, evolved by some kind of mutation, and stuck around because (by mere chance) it coincided with physical causality and it gave our ancestors a substantial edge for survival. By this account, causality in the mind might be *the* reason human intelligence became so superior. Being able to process causality is not only a hallmark of human intelligence, but it is its most fundamental and most important

⁵Which is actually Kantian.

component.

But sticking with this argument one might ask a rather basic but easy to miss question. Why is it helpful to understand causality? Take as an example a strong wing knocking down a tree. It seems that we need causality to understand it, but actually we need nothing more than instincts to avoid this dangerous situation. In fact, most animals do exactly this: they get startled and run. They do not form theories and reason about them. This does not mean having a mental capacity to process causality is useless, far from it. Take for example a pack of rats that is being poisoned. Usually if the pack is big enough, the rats somehow survive and learn to avoid the poison. The usual explanation is an evolutionary one: most rats die off, but some have a mutation that enables them to resist that particular poison. But if rats, like people, are able to understand causality, this would enable them to learn and pass on knowledge on the danger of that poison. This approach can be called “cultural”, lacking a better word. Cultural adaptation needs causality to understand and pass on information, and the advantage over evolution is huge: there is no need for most of the population to die while waiting for a particular mutation to develop. In fact, the mutation needed could be aptly called “extraordinary”, while the understanding of causality has to be “ordinary”, i.e. possessed by the majority of the pack for cultural adaptation to work. This simple mental faculty helps in all kinds of situations, and offers great advantages over evolution, since it is applicable to a large number of different dangers, whereas mutations brought on by evolution protect against a single danger.

Following this argument, if the rats evolved immunity to a given poison, they should still be vulnerable to a different poison administered in the same way. If cultural adaptation is the name of the game, the situation is much harder for the exterminator, since there is no telling what the rats have learned to avoid. Just the concrete delivery method? Anything that smells “funny”? Anything that has residual human smell? Anything that is not common in their environment? Even though it is hard to pinpoint it, one thing is certain: the very thing that makes the situation difficult for the exterminator is exactly what gives the rats a huge advantage. In a sense, the rats using causality to achieve cultural adaptation trim their own mental search space of possibilities needed to be explored, and “decide” to forgo some of those. By using a clever and causally deeper approach (e.g. to avoid everything that is usually not in their environment) combined with the sheer number of possible approaches, easily gives them a simple *modus operandi*, while at the same time disabling almost any trimming of the solution search space for the exterminator. None of this would happen if all there was here was plain biological evolution. Of course, what is actually the case in such a scenario (pure evolution or cultural adaptation) can in fact be determined experimentally, at least in theory.

Search space and intelligence

An interesting question arises here. If in fact intelligence is a search, and having a faculty for processing causality helps to reshape the search space, is there anything else that might help? Reality in itself is vast and also broken up. Sometimes we have very similar yet distinct possibilities, and sometimes a vast array of seemingly different possibilities are all the same. To combat this, our mind has a distinctive faculty which can use previous experience to model an alternative, more simple reality, which can then be processed more easily. This faculty of the mind also fills the conceptual empty space by producing classifications and interpolations, also known as presumptions. By doing so, search becomes easier, in part because the search space can be deemed conceptually “completed”, and in part because the complexities of reality get replaced by a more manageable alternative, consisting of N categories, each with its own hypothesized development stretching in the future, which is used as such, and amended should the predicted future experience be out of tune with the current reality. This faculty is our imagination, and it has been neglected by most philosophers, with the notable exception of Carl Gustav Jung (Jung, 2016).

The whole idea of intelligence as search is not a new one, it is in fact the oldest idea out there. It was the one used by Herbert Simon and Alan Newell to create the first AI system in history, the Logic Theorist. By using the intelligence-as-search and a bit of dualism for its philosophical foundation, Herber Simon (Crevier, 1993) claimed that he and Alan Newell invented a “thinking machine” (the Logic Theorist), and thus “solved the venerable mind-body problem”.

The interesting point here is not whether they invented a thinking machine (and they were certainly not the first ones to lay the foundations for one, e.g., Turing or Pitts and McCulloch beat them to it), but the interesting thing was the fact that the found it natural to think that the proof of existence of a “thinking machine” actually solved the mind-body problem, as if it was crystal clear that “thinking” is strictly mental, and “machine” was strictly physical, where “thinking” could not be viewed as a process which could be realized over a material substrate,⁶ and, what is perhaps even more important, that proving theorems, which is the only “thinking” the Logic Theorist does and which is a very special kind of thinking, characterized by very mechanistic ratiocination and step by step procedure is enough to claim they captured thinking in general, and, moreover, intelligence in general, and the human mind in general. Theorem proving is certainly not only not representative, but not even stereotypical of

⁶Let us put aside that the brain is one such substrate for now.

thinking in general. Also, there is much more to intelligence than just thinking, and much more in the human mind than just intelligence. This points to a highly reductive version of dualism that they actually advocated.

What is even weirder is the fact that not only is this approach too reductive for capturing the human mind, but it is also too general to provide any useful insight once complexity hits the fan. It certainly seems that searching as an approach to intelligence is too general and too mechanical. It is too general since, by definition, every problem in the world is essentially a search problem: given a problem, one *searches* for its solution. Since any conceivable problem is a prompt for a search, and therefore, any problem-solving is in essence searching for a solution (among all possibilities). The second problem is that it is too mechanical. Here the issue is with thorough search, which explores all possibilities. It is very hard to see any kind of intelligence hidden behind such a simple exhaustive search. Imagine asking a person what is 34×67 . And imagine that in return all the person does is give a list of answers, starting from 1, continuing to 2 and so on. This seems to be the contrary of intelligence, since not only would we not call such a search intelligent, but we *would* call a restriction of such a search intelligent. Imagine that faced with the same question the person says that the product of two digit numbers has to have at least three digits, and then starts again 100, 101, etc. We would be prone to call it more intelligent than the first attempt *because* it trims the search space. An exhaustive search always finds a solution if a solution exists, but any other kind of search, even the ones that do not find any solution, seems to be more aptly described as “intelligent” if the criteria for trimming can be described as “smart”.

Then there is also a combinatorial problem with exhaustive search. Any search produces a search tree, whose nodes are possible actions. When searching exhaustively for a solution, more often than not, a sequence of actions which is explored is abandoned and a new line of actions is explored. This is called “backtracking”, and it is often at least of exponential complexity. Clever ways have been found for particular problems, but no universal way has been found. This is the essence of the famous **P** vs **NP** problem defined in 1971 by Cook (Cook, 1971). Here we are not interested in this problem, but the idea is that anything that goes against exhaustive search is considered “smart”.

Strategies that make the problem-solving intelligent tend to involve creativity, and to a degree, welcome imprecision. We, as kids, collected stickers of footballers, usually issued during FIFA World Cups. Collecting them all was a combinatorial nightmare. There were around 200 stickers in total, each displaying the face of a footballer. When you got the first packet of 10, you got all new ones, no duplicates. Once you get five packs or so, duplicates start coming along. And the more stickers you had, the greater the chance to get a duplicate. In fact the last stickers (say the last 10) were so hard to find

that you had to buy ten or twenty packs to get just one which you are missing. And as you near completion, it gets almost impossibly hard to complete the sticker album. Trading doubles with friends helps, but not a lot. If you consider the input to be the n -th sticker, and the output the number of packets you need to obtain it, for the first one the result is 0.1 (all ten stickers are new), but as it progresses it is not linear, it is in fact exponential. It takes a lot of packets to find all 200.

Faced with such a computational nightmare, we devised a somewhat clever idea. Instead of insisting that only a sticker with Roberto Baggio can be placed on the space for Roberto Baggio, any footballer with dark short curly hair and no beard or mustache can be put on the album slot normally reserved only for Roberto Baggio. By accepting that the solution is not *exact* but only *approximate*, we trimmed the search space considerably. In fact, if you lower the similarity requirement enough, you will be able to put almost any sticker on any album slot, thereby lowering the complexity even to linear. There was also a variant in which we used the sticker number similarity, i.e. in the place of sticker no. 12 you could place 11, 13, but also 21 and 112, which we thought as less clever than the physical similarity criteria.

Back to the future

Let us go back to 1956, to the Dartmouth summer project on artificial intelligence. John McCarthy wrote a manifesto, and its second sentence can be taken to be the maxim defining AI as such, and in that sense, still relevant as ever (McCarthy et al., 1996):

The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.

This manifesto has two parts, defining two equally important tasks for AI. The first task is to *precisely define* every aspect of learning and other forms of intelligence, and the second is to use that definition to produce a technological replica. The importance of the first segment cannot be overstated, for the higher its precision, the easier it will be to implement it in a machine. This means a real and intrinsic collaboration between the people who provide these definitions, i.e., philosophers, and the people who will implement the given notions, i.e. engineers and computer scientists. Not only does this maxim delineate the respective tasks for these two endeavors, but it places clear methodological requirements on their communication. In order to understand each other, the philosopher must deliver his definitions with formal and mathematical clarity, and this emboldens the importance and position of logic within philosophy. It makes sense to expand the curriculum in logic to include parts of what is traditionally thought to be computer science topics, so that, by seeing how her

definitions will be used, the philosopher has an easier time formulating them in a clear and precise fashion, so that the computer scientist can hit the ground running.

Herbert Simon was the first to explore in his book (Simon, 1996) the idea of a new science, artificial intelligence, as a reproductive science. In this view, to understand a topic is no longer considered synonymous with analysis, but with synthesis, reproduction. Even though this opens a new paradigm in the whole of science, there is a more pressing point: every reproduction must begin with the understanding of the subject matter, and in artificial intelligence the understanding part stood squarely in philosophy. As the science evolved, the need for including related sciences such as linguistics and psychology grew and culminated when Christopher Lounget-Higgins defined cognitive science which would go on and be slightly redefined as consisting out of six equally important sciences, namely philosophy, psychology, linguistics, anthropology, neuroscience and artificial intelligence.

Analysis is a valid method in most of these sciences, but only one of them deals with definitions: philosophy. The analyses serve as a basis for definitions, and converge to bring the elements needed to formulate it in the manner that McCarthy postulated. Four of the sciences analyze and funnel their results into philosophy, which produces definitions, and then they are passed to artificial intelligence for implementation. Once the definition, perhaps an imperfect but still a precise one, is in place the reproductive part, the programming can start. But what does it mean to define the mind? Or some subroutine of the mind? Could we hope to proceed to build the artificial mind as such and treat partial results, modules that implement faculties as parts of the mind? Or is the mind in a sense “atomic”? In this view, we can only hope to build the whole artificial mind and the modules we have so far, which we are so tempted to consider as real partial results in artificial intelligence are plain software as close to the human mind as simple HTML web pages.

(Re)defining the mind

It is obvious that we are nearing the point where we shall need a working definition of the mind. But what would this definition look like? Tossing aside the elegant escape in claiming the undefinability of the mind, two basic approaches come to mind. The first one would be materialistic. In essence it would be a technical specification of how to connect neurons to get mental processes, similar to a set of building instructions. The second approach is the classic dualist approach. We will define the mind using mental terms, and by abstaining from neurological components. Should we need the neurological parts, we will never consider the exact technical specifications but rather the abstractions such as recurrent connections between neurons. In this

way we will stay within the mental part of the dualist language. Oddly enough such an approach is easy to understand and actually easier to reimplement than the materialist one. Even though it sounds in theory easier to implement an artificial mind using neuron-by-neuron specification, this approach would require artificial neurons to be the *exact* artificial equivalent of real neurons. Here teleology creeps in and drastically complicates things. Consider an artificial heart. What makes an artificial heart an *exact artificial equivalent* to a real heart? There is a very simple answer here: because it is made to fulfill the same purpose a real heart does. It pumps blood in *the same way* a real heart does, and this is its essence. How it looks or how it is powered or installed is far less important.

The problem with the materialist approach of using exact artificial equivalents of neurons which are to be connected with a neuron-by-neuron specification is that, unlike with the artificial heart, we do not really know their purpose. All other organs in the human body cater to a specific need of the brain. The heart provides oxygen, the skin provides both a barrier and sensory input, but they are in a sense like machines that the brain employs. The brain itself is in charge, and because of that fact it is prohibitively hard to specify its “purpose”, and therefore impossible to recreate that purpose in component neurons. The dualist approach seems to circumvent these issues.

Before continuing to dualist definitions of the mind, Let us return to the interesting danger we mentioned earlier. There is an old philosophical discussion on the nature of freedom. Are we determined by physical laws, as nature is, and our freedom is an illusion, or are we really free? The advent of quantum physics saw a large number of subintelligent and unimaginative philosophers flock to a simple novel argument: quantum physics says nature is not determined, and therefore we are free. This argument is in a way similar to the one which says nature is determined and we are not free. They both consider nature and the human mind to be similar in this regard. The opposite view is the one that combines determinism in nature and freedom in the mind, and this approach is aptly called “compatibilism”. Compatibilism is not easy to argue for, but it is important since its arguments show how a free mind might have come into existence in fully determined neurons. The interesting danger in our analysis is in fact a compatibilist argument. Humans can understand dualist concepts well. They are easy to process and quick to internalize. A neuron-by-neuron specification is not only hard to process for humans, but it bears no meaning. Specifying a list of neural activations is certainly not the same as saying to someone “I am happy”, and in a sense it is also less precise. It does a worse job at presenting a mental state. What if the neuron-by-neuron specification is so much more complex that humans cannot even interpret it? The dualist language then is the only way to go forward, even though it might precisely map to a neuron-by-neuron specification. The

impossibility of processing such a specification makes the dualist description substantially different, and it becomes more than just something that needs to be rewritten more precisely, because the rewriting would result in an uncheckable set of connections, which cannot even in theory be the same as a simple dualist phrase which describes it.

So, let us take the easy road, the dualist lingo. A first attempt of defining the mind could be to state that the mind is what controls the body. A huge problem occurs right away. First we can ask, does it exist in the same way as the body or does it exist in a different way? If it exists in the same way, it seems that the mind is equal to the brain (or some part of it), and then we are required to explain mental faculties in terms of brain processes. Anything less is neither scientific nor intellectually honest. Describing the mind and its processes with specialized lingo (such as the one commonly employed by psychologists, is not only metaphorical, but it actually describes nothing real--for in the "real" world there are only neurons and synapses. Anything less than that is exactly like saying that emotions exist and "live" in the heart. Asides from pushing psychology, linguistics, sociology in the realm of pseudosciences, this approach has a second completely unacceptable consequence, and that is that we know nothing about the mind which we cannot describe completely in terms of synapses and neurons. And the emphasis here is on "completely". This means we literally know nothing about the human mind--it is as alien to us as the minds of insects or octopuses. If on the other hand the mind exists in the brain but in a different way than the brain, we can formulate a modified version of Plato's third person argument #: if the mind and the brain are substantially different, and yet are linked together, does that link share the substance of the mind or the brain? Let us call the substance of the mind "ideal substance", and the substance of the brain "material substance". If the link is of the material substance, how does it link to the mind? If it is of the ideal type, same as the mind, how does it connect to the material brain? The whole idea of introducing an additional link solves nothing.

Conclusion

Navigating the labyrinthine paths of defining the mind reveals a terrain fraught with philosophical and scientific challenges. From McCarthy's foundational conjecture that every facet of intelligence can be precisely described and simulated, to Simon's synthesis of AI as a reproductive science, the quest for a definition of the mind unfolds as a dialectic between materialistic and dualist perspectives.

The materialist approach, akin to architectural blueprints for connecting neurons, strives for technical precision but falters on the enigmatic purpose of neural components. In contrast, the dualist stance, embracing mental terms divorced from exact neural

specifications, offers clarity and intuitive grasp, albeit with the risk of abstraction.

These contrasting perspectives underscore a deeper quandary: whether the mind is reducible to brain processes or possesses an ineffable essence beyond neuronal firings. The dualist path invites us with its intuitive appeal, suggesting the mind as the conductor of bodily functions transcends mere neural mechanics. Yet, it requires rigorous scrutiny to avoid marginalizing psychology and linguistics as pseudosciences and to maintain intellectual integrity in bridging mental and material realms.

Ultimately, whether the mind emerges from neural networks or exists as a distinct entity intertwined with the brain remains an open question. The journey toward defining the mind necessitates a synthesis that accommodates the rigor of neuroscience while embracing the complexity of mental experience. Only then can we hope to unravel the elusive nature of the mind, a challenge as profound today as it was during the Dartmouth AI summer project of 1956.

References

- Aristotle. *The Nicomachean Ethics*. Bartlett RC, Collins SD, trans. & eds. University of Chicago Press; 2012.
- Ashby R. *Introduction to Cybernetics*. Martino Fine Books; 2015. Original work published 1956.
- Berkeley G. *A Treatise Concerning the Principles of Human Knowledge*. Turbayne CM, ed. Forgotten Books; 1957.
- Cassirer E. *An Essay on Man: An Introduction to a Philosophy of Human Culture*. Yale University Press; 1944.
- Cook S. The complexity of theorem proving procedures. In: *Proceedings of the Third Annual ACM Symposium on Theory of Computing*. ACM; 1971:151–158.
- Crevier D. *AI: The Tumultuous Search for Artificial Intelligence*. BasicBooks; 1993.
- Descartes R. *Discourse on Method and Meditations on First Philosophy*. Hackett Publishing Company; 1998.
- Geffer A. *The man who tried to redeem the world with logic*. Nautilus. 2015.
- Gettier E. Is justified true belief knowledge? *Analysis*. 1963;23:121–123.
- Greif H, Šekrst K. The origin of adaptation, effective procedures, and the reality of mechanism. In: *IACAP 2024*; 2024.
- Hume D. *A Treatise of Human Nature*. 1739-1740. Accessed October 13, 2024.
- Jung CG. *Psychological Types*. Martino Fine Books; 2016. Original work published 1924.
- Kant I. *Critique of Pure Reason*. Guyer P, Wood A, trans. & eds. Cambridge University Press; 1997.
- Kripke S. *Wittgenstein on Rules and Private Language*. John Wiley & Sons; 1984.
- Locke J. *An Essay Concerning Human Understanding*. In: *The Works of John Locke in Nine Volumes*. 12th ed. Vol 1. Rivington; 1824.
- McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*.

- 2006;27(4):12.
- McCulloch WS, Pitts W. A logical calculus of ideas immanent in nervous activity. *Bull Math Biophys.* 1943;5:115–133.
- Minsky M. *Society of Mind.* Simon & Schuster; 1988.
- Pace NR. The universal nature of biochemistry. *Proc Natl Acad Sci U S A.* 2001;98(3):805–808.
- Perkov T. The McCulloch–Pitts paper from the perspective of mathematical logic. In: Skansi S, ed. *Guide to Deep Learning Basics.* Springer; 2020:7–12.
- Plato. *Phaedo.* Fowler HN, trans. In: Lamb WRM, ed. *Plato in Twelve Volumes.* Vol 1. Harvard University Press & William Heinemann Ltd; 1966. Original work published 1925.
- Plato. *Parmenides.* Scolnicov S, trans. & commentary. University of California Press; 2003.
- Ryle G. *The Concept of Mind.* University of Chicago Press; 2000. Original work published 1949.
- Quine WV. *From a Logical Point of View: Nine Logico-Philosophical Essays.* Harvard University Press; 1980.
- Shannon CE. A symbolic analysis of relay and switching circuits. *Trans Am Inst Electr Eng.* 1938;57(12):713–723.
- Simon HA. *The Sciences of the Artificial.* 3rd ed. MIT Press; 1996.
- Skansi S, Šekrst K. The role of process ontology in cybernetics. *Synth Philos.* 2021;36(2):461–469.
- Turing AM. On computable numbers, with an application to the Entscheidungsproblem. *Proc Lond Math Soc.* 1937;2(42):230–265.
- Wittgenstein L. *Philosophical Investigations.* 50th anniversary ed. Blackwell Publishers; 2001.