The Evolution of Reason
Have We Become Rational?

Stephen N. Lyle

May 18, 2022

This will not be about any recent evolution of human reason. We adopt here a biological point of view of evolution and reason, and hence of ourselves, the human species. What motivates this? Well, it is often said that what distinguishes us most from other animals is our ability to reason, our faculty of reason. But this raises a question: why did our species evolve this trait? What advantage did it give us, such that the genes which predispose us to reason would proliferate in our species?

To go further, we need to decide what we mean by reasoning. In his recent book *Rationality*, the Canadian cognitive psychologist Steven Pinker gives us this idea: reasoning involves using knowledge to achieve an objective [1]. But this kind of reasoning would hardly distinguish us from other animals. All animals collect information from their environment with their various senses, process it in their body and in particular in their brain, if they have one, and act accordingly. In this sense, one could say that all information processing is a kind of reasoning, an idea that will be developed here. And note that animals, including ourselves, are not necessarily aware of this kind of information processing, so this may well be a kind of unconscious reasoning, another idea that will be developed later on.

If we want a notion of reasoning that would really distinguish us from other animals, it could be something like this: to use logic to solve problems and explain ourselves clearly. This seems to involve language, which largely rules out most other animals, although one should remain open. So let us consider briefly what logic is. Here is a short argument in generic form:

If A, then B
Not B
Hence, not A

There are two premisses and a conclusion. A and B are propositions, that is, things one could say about what is happening in one's environment. This can be illustrated by an example. If it rains for a whole week, the stream overflows its banks. But the stream has not overflowed its banks. Therefore, it has not rained for a whole week. This kind of short argument is called a syllogism and it even has a name: *modus tollens*. This is in Latin, which shows that humans have known about syllogisms for a long time!

Here is another syllogism, called *modus ponens*:

If A, then B
A
Hence, B

Once again, there are two premisses and a conclusion. We can illustrate with the same example. If it rains for a whole week, the stream overflows its banks. It has in fact rained for a whole week. Hence, the stream will be overflowing its banks right now.

These two short arguments are clearly *valid*, whatever their content as specified by the propositions A and B. They are valid by their form alone. Validity

Figure 1: The greater kudu (*Tragelaphus strepsiceros*), a woodland antelope found throughout eastern and southern Africa. Credit: L0k1m0nk33 at English Wikipedia

is a precise technical term for logicians: an argument is valid if, whenever the premisses are true, the conclusion is necessarily true. And logic is the study of the form of valid arguments, and by extension, the study of the form of invalid arguments.

So much for logic. But are we logical? And if we are, have we always been so? What about our distant ancestors *Homo sapiens*, a hundred thousand or two hundred thousand years ago? How could we find out? Figure 1 shows a kudu, an animal that lives in the Kalahari Desert in southern Africa. This animal is hunted, sometimes stalked over great distances, by the San, hunter–gatherers living in the same region. For their part, the San have long been stalked by Western anthropologists, precisely because we think such people may live the way our distant ancestors lived, hunting and gathering, a hundred thousand or two hundred thousand years ago. And of course, one of the things that interests us is to find out whether they are logical! Here is just one example to confirm that they most certainly are logical, given by Pinker in his book [1]:

> The kudu can be caught in the dry season as it tires easily when it has to run on loose sand
>
> It's the dry season and the animal that left these traces is a kudu
>
> As a consequence, we should stalk this animal

There's no real surprise here. Pinker gives many other examples of the San's rationality. It turns out that their rationality is the same as ours. And their irrationality, too. These things are clearly going to be universals, common to our whole species, and this throughout its existence.

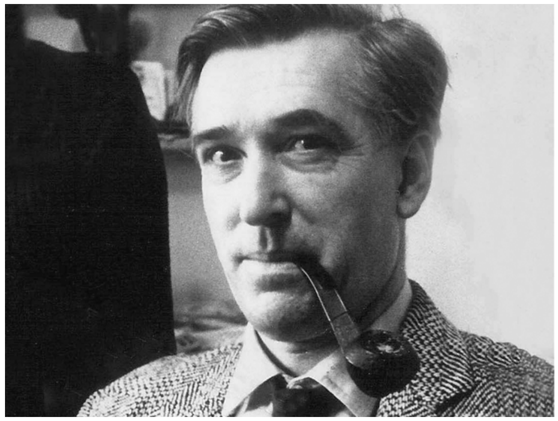

Figure 2: Peter Cathcart Wason (22 April 1924–17 April 2003) was a cognitive psychologist at University College, London who pioneered the psychology of reasoning. He put forward explanations as to why people make certain consistent mistakes in logical reasoning. He designed problems and tests to demonstrate these processes, such as the Wason selection task. Figure credit: https://medium.com/five-guys-facts/the-wason-selection-task-4521cd54cc7b

But are we always logical? Figure 2 shows the British cognitive psychologist Peter Wason, clearly inspired by Sherlock Holmes, who introduced a very clever and revealing test of logic in the 1960s. The Wason test seems very simple (see Fig. 3). We are presented with four cards. On each there is a letter on one side and a number on the other, and we are told that the cards should obey the following rule: if there is a vowel on one side, there must be an even number on the other. So which cards must we turn over to check that the rule has been respected? Many people reason like this. The first card shows a vowel, so we must check that there is indeed an even number on the other side. The second shows a consonant, and since the rule doesn't mention consonants, no need to check the other side. The third shows an even number, and since the rule mentions even numbers, we need to check that there's a vowel on the other side. The last card shows an odd number, not mentioned by the rule, hence no need to check.

But that is wrong. Of course, some people don't make the mistake with the last two cards. Logicians, for example. We don't need to turn the third card, because whatever is on the other side, consonant or vowel, the rule is respected. However, if the last card had a vowel on the other side, the rule would not have been respected, so we do in fact need to check that one.

Well, that shows that many people don't apply those simple valid logical arguments correctly. But look at the four Wason cards in Fig. 4. This time each card has someone's age on one side and what they are drinking on the other. And this time the rule is of course that one must be at least eighteen to drink alcohol. Which cards must be turned over now? This is very easy and

Figure 3: Four of Wason's cards. On one side of each there is a letter and on the other a number. And there is a rule: if a card has a vowel on one side, it has an even number on the other. Which cards must we turn over to see whether the rule is observed?

Figure 4: Four more Wason cards. On one side of each there is someone's age and on the other a picture of what they are drinking. Once again, there is a rule: you must be at least eighteen years old to be allowed to drink alcohol. Which cards must we turn over to see whether the rule is observed?

nobody, absolutely nobody makes a mistake here. The first card must be turned over to check that this sixteen year old is not drinking alcohol. We don't need to check the second, because at twenty-five, the person can drink what they like. The third person is drinking soda, so it doesn't matter how old they are, but the fourth is drinking beer, so we do need to check.

So what? The point is that, logically speaking, the second test is *exactly* the same as the first. So how come so many people make mistakes with the first version, and no one with the second? The answer is that the first test is abstract while the second deals with deontic conditions, that's to say, conditions relating to duty and obligation. The fact is that we are extremely good at holding other people responsible for their acts, and equally good at finding our way in a world in which others will hold us responsible for our own. And that is because we live in a cooperative society. Indeed, humans are the world champion cooperators, far more so than any other species. We can cooperate on anything and everything. Other animals like wolves, chimpanzees, bees, and ants may live in communities and work together in various ways, but humans are continually inventing new ways of doing things together for better result.

But it comes at a price, because we need to know whom we can trust and we need others at least to tend to be trustworthy. In other words, we need a system of ethics. The key here is morality. So look at this: we set out to make a cognitive theory of reasoning, and here we are talking about morality. This is no accident. The consensus in cognitive psychology today considers that reasoning evolved in humans for the purposes of moralising, to make it possible to live in a predominantly cooperative society. The benefits for genes leading to such behaviour are legion and they have proliferated. They are almost everywhere today in human society. It is sometimes said that there has been a kind of self-domestication over the last few tens of thousands of years, i.e., a coevolution of genes and cultural practices that has led to certain genes, these genes for cooperation, becoming dominant. We shall return to this important theme later.

But following Pinker in his latest book, let us just ask what kind of content A and B in our syllogisms would transform us temporarily into reliable logicians. In his view, it is clear that "they must embody the kinds of logical challenges that we became attuned to as we developed into adults and perhaps when we evolved into humans" [1, p. 15]. We have seen that the San hunter–gatherers correctly apply syllogistic reasoning when they hunt together. It would be disastrous if they didn't, given the importance of the task. But when a problem is not really important for us, like the first, totally abstract Wason selection task shown in Fig. 3, we tend to go for the easy answer, even if we know very well it may be incorrect because we have hardly invested the mental effort needed to sure. And here is another important point: labouring over logical conundrums is a costly exercise, in terms of both energy and time. Try this, for example, also from Pinker [1]:

> Eight printers take eight minutes to print eight copies of a document.
>
> How long will twenty-four printers take to print twenty-four copies of the same document?

No, it's not twenty-four minutes! But hopefully, in a real situation with the printers sitting in front of you, you would get the correct answer right away.

The Mind

Back at the beginning, we said we would adopt a biological point of view of reason, and hence of ourselves. But that is not how we usually understand ourselves when we introspect. Of course, I can *see* my rather animal body, but when I introspect, I find ... me! My mind. The self. And of course, it is the mind or the self that thinks and reasons, is it not? This is a completely natural dualism: mind and body, or body and soul. But what is the mind? How does that work? And how could such a thing have evolved?

To tackle these questions, it is interesting to look to the two main paradigms of artificial intelligence. For it should not be forgotten that one of the main aims of research in artificial intelligence is to understand natural intelligence—our own intelligence. At the time when computers were being developed, since the Second World War, and particularly in the 1960s and thereafter, it was natural to wonder whether our own brains might function like these new computing machines. After all, computers process information, and they are logical, even exclusively so. This led to an AI paradigm that is sometimes rather pejoratively referred to as good old-fashioned artificial intelligence or GOFAI.

This paradigm maintains the dualism we find in our ordinary perspective of ourselves. There is hardware and software. A GOFAI robot first plans and calculates, then acts. But this requires an enormous amount of memory and considerable computing power. One reason is that it comes up against what is known as the *frame problem*. This is basically the problem of representing the robot's environment and the changes that have to be made in that representation whenever the robot does something. Determining what an action changes and what it doesn't change is no simple matter, especially for systems that can only act by applying pre-installed principles. For a GOFAI system, all the possible consequences of an action must be programmed into the software using algorithms the computer can execute. You have to make a long list of all the changes necessary in your representation of the world for each possible action, and an even longer list of all the changes that you don't need to make.

Here is a first lesson from this research in artificial intelligence: there are things that are very easy for us—distinguishing an apple and a pear visually, walking, catching a flying object, etc.—but that doesn't mean that they are simple. They sometimes turn out to be very, very difficult to program into a robot. We shall return to this point with a solution in part later on. But for the moment, the question it raises is this: do human beings really work like this? And if not, how do we avoid having to calculate everything before we act?

Here is another thing that is relevant to this issue. It seems that very young children, even as young as three or four months, have expectations about the behaviour of inert matter. In fact, they seem to apply certain simple principles of physics, sometimes called folk physics. We know this even though these children cannot yet tell us, because they are too young to talk. The trick is to watch their eyes when they are looking at different scenarios simulated on a screen, some being physically possible and others physically impossible, to see if they do a double take in the latter case (see [2] for an example). This process, known as eye tracking, measures either the point of gaze (where the child is looking) or the motion of an eye relative to the head. Eye trackers are widely used in research on the visual system, but also in psychology and psycholinguistics, and many other applications. Without going into the details of the principles of folk physics that have been established, some examples are as follows:

- Principle of solidity. One solid object cannot pass through another.
- Object permanence. An object continues to exist even when it passes behind another and is no longer in view.
- Principle of cohesion. A moving object maintains its connectedness and

boundaries.

A general review can be found in [3].

What interests us here is the existence of physical principles in the baby's mind, principles that are quite possibly innate. These infants clearly draw inferences about things they are watching. Now, the basic principles of folk physics could be represented explicitly and applied in the same way as they are applied in computers, and as one might try to do if one wished to program a robot. This is the computational approach of GOFAI, which assumes that a cognitive system can only act in accordance with principles if these principles have been explicitly represented within the system in a suitable symbolic form.

This was taken to be the founding principle of GOFAI, known as the *physical symbol hypothesis*: rules and principles are explicitly represented in symbolic form in the brain. One could say something like this: syntax tracks semantics. But that is an action-packed statement that needs, as the philosophers would say, a certain amount of unpacking. The idea is basically that it is the very structure of the sequence of symbols used that contains the meaning of that sequence. In the above list, the principles are explicitly represented in the English language. This can be translated into a computer language expressed in the well known binary alphabet of zeros and ones, and this in turn can be translated into a configuration, let's say of atoms, inside the computer, where that configuration contains the meaning of the principle in the very structure of their arrangement.

However, it turns out that one can develop what are known as *artificial neural networks* (ANNs) capable of simulating the behaviour of human babies in these experiments without explicitly encoding any rule or principle. So here we have a proof of concept that the computational approach is *not* the only option. Something that looks like rule-based behaviour can emerge from a system that exploits no explicitly represented rule. So there can be a form of information processing that does not involve the kind of manipulation of symbols governed by rules put forward by the physical symbol hypothesis. Indeed, no principle is stored as such in an artificial neural network.

The problem with GOFAI is that it is an intellectual top-down approach to AI. All the intelligence is in the programming. And it is too slow. We're just lucky that computers calculate so fast, because they have to optimise and plan everything in advance. Imagine the calculations involved in getting a GOFAI robot to find its way through a forest, avoiding the many obstacles by applying preprogrammed principles.

Just a word about these ANNs. The approach here is really quite the opposite to what happens in GOFAI. In that case, we started with the computer and asked whether our brains might work like that. With ANNs, we start with the brain and make highly simplified simulations of real neural networks, i.e., we actually look at certain features of neurons and imitate those in the ANN. So an ANN is an information processor with some of the characteristics of the brain, but far, far fewer neurons. And it can learn, in a certain sense. Hence the name 'deep learning' for this AI paradigm. That is, the ANN is presented with lots of data concerning the question it is supposed to be able to answer, in situations where the answer to that question is known. When it gives an inaccurate answer, its parameters are adjusted in a well defined way before presenting it with more data, and so on and so forth until it always, or almost always, gives the right answer. A good introduction to artificial intelligence, with no hype, is the excellent book by the AI researcher Melanie Mitchell [4].

Thinking

A moment ago, we said that it was the mind or the self that thinks, or indeed reasons. But we have just been talking about artificial *neurons*, inspired by real neurons in our brain. A neuron is a cell and there are something like 80 billion of them in a human brain, along with many other kinds of cell. We know that information passes along these cells and also between them, with each neuron being connected to on average 7000 other neurons. And now we know that, imitating a few simple features of these cells, we can build up networks of artificial neurons that process information, even seem to apply principles that are nowhere encoded symbolically within them. This strongly suggests that it is somehow the neurons and their entourage of other cells that think and reason. Indeed, there is every reason to think that the self is just a representation generated by the brain to represent the individual in which it resides, very likely for the purposes of that individual's interactions with other humans, in fact, with other selves.

The natural dualism of mind and body is then just an impression we have. And if we hope to understand ourselves from the biological point of view, we really need to get past the story told by introspection and adopt a cognitive—or neuronal—perspective of the individual, to use a turn of phrase promoted by the linguist Ray Jackendoff in his excellent book *A User's Guide to Thought and Meaning* [5]. Basically, this means starting out with what we can observe—matter, its configurations, and its dynamics—and somehow arriving at the self without adding any other ingredient. This is the aim of the naturalist philosopher.

There is a well known anecdote about J.M. Keynes, the famous British economist. Bertrand Russell said Keynes was the most intelligent person he had ever met, which is not a bad reference. Someone once asked Keynes if he thought in words or in pictures, an age-old philosophical conundrum. Keynes thought for a moment, then said: "I think in thoughts." This may not seem like a very useful remark. However, if he was suggesting that we think neither in words nor in pictures, then he was close to the position of many neuroscientists today. For in the cognitive perspect, it is generally agreed that *all our thoughts are unconscious*. This may sound mad, but consider this. When you have to communicate some rather subtle idea to someone else, don't you find sometimes that you haven't said exactly what you intended? It wasn't exactly what you mean? Then you reformulate, maybe several times, adjusting what you say until at some point it sounds right. It finally corresponds to your idea. Think of what happens when you have to write a rather difficult letter, or an essay. How many times do you end up crossing something out or adjusting it?

So why do we say things that don't mean what we intended? Jackendoff and others will say it's because that meaning is actually hidden in our unconscious—this is the *unconscious meaning hypothesis* [5]. What we ordinarily call thinking is merely a linguistic take on our actual unconscious thinking, a symbolic representation in our consciousness using the symbolism of language to represent thoughts and meanings that actually lie deep within the configurations and dynamics of our neurons. 'Conscious thought' or 'conscious reflection' is then just a (logical?) sequence of linguistic takes on our actual thinking, which for its part is carried out unknown to us—that is, unknown to our introspection—by our neurons.

This brings us back to good old-fashioned AI, which is based on the idea that we think the way we think we think, in our ordinary perspective of ourselves. We think we think in logical sequences of sentences—or propositions, to use a heavily connotated philosophers' term—i.e., sequences of symbols like those we input into our computers and which *they* can manipulate using algorithms. But cognitively, dropping for a moment the claims of our introspecting self, thinking may not actually be a matter of applying logic to premises and principles. We know from the development of ANNs that there is absolutely no need for premises and principles to be coded somewhere in our brains. In the ordinary perspective given to us by introspection, we always have the impression that information must be coded somewhere because our (introspection's) only take on information, and in particular its meaning, is via the 'little voice' we hear chattering away almost permanently inside our heads. But it could be that these thoughts and meanings are only encoded by language for the purposes of communication, with others or indeed with ourselves. And in reality, as we have seen, we often have difficulty reasoning in a strictly logical way, something our computers do marvellously well by design.

Evolution and the Modular Mind

We have just been talking about neurons and the brain. It seems rather clear that it is in fact these purely biological entities that do our thinking and reason for us. So we can now return to an earlier question and rephrase it: how does *our brain* avoid the need to calculate everything before it instructs us to act? To approach this, it is useful to ask *how* we humans were designed?[1] But we know the answer: we have been slowly cobbled together by evolution. To quote the Scottish biologist D'Arcy Wentworth Thompson: "Everything is the way it is because it got that way." So let us ask how the human mind, or any other animal's mind, could have evolved.

An important idea in evolutionary psychology, i.e., in the evolutionary way of

[1] This word is used in the sense proposed by the philosopher Daniel Dennett to refer to design by genetic mutation and natural selection, where there is no designer (see, for example, [8]).

understanding cognitive activity, is the modular theory of mind. In a word, this is the idea that the mind is not unitary and based on principles, but modular and opportunistic, an assemblage of bits and pieces that evolution has "put together" to deal with specific problems faced by our ancestors—not just our human ancestors, but all our ancestors, as we shall see.[2] In this view, the animal mind is made up of modules which draw inferences, each in its own specific domain, and this over a whole range of widely varying domains. A module processes information, hence reasons in the broad sense mentioned at the beginning, and the animal acts in consequence. Each species evolves, building module after module, giving it different abilities in its interactions with the world.

This will be illustrated in a moment, but first a little history. The idea of a modular mind could at a push be traced back to the largely discredited idea of phrenology: studying the size and shape of the cranium was supposed to tell us about the character and mental abilities of a person. The idea was rehabilitated rather differently by the philosopher Jerry Fodor in the 1970s [6], relaxing the idea that modules must necessarily be localised, i.e., correspond to specific regions of the brain. His theory was that certain specific cognitive activities might be handled by a specially designed module—each module would work in a rather autonomous way on the specific problem it had evolved to solve, e.g., the perception of colours, analysis of shapes, analysis of relations in space, visual guidance of body movements, face recognition, grammatical analysis of spoken language, voice recognition, and many, many others. But in his view, modules would treat only specific problems, each involving only a certain type of information. This was referred to as information encapsulation.

All these considerations concern the question of the architecture of the mind: what different kinds of information processing go on inside it, in what different kinds of structure, and with what connections between those different kinds of structure? But one thing is clear: modules are now generally accepted to play a key role in this. And the modules in question are Darwinian in the sense described above. The idea came to the fore in 1987 in work by the American anthropologist John Tooby and the American psychologist Leda Cosmides [7]. Unfortunately, we can only skate over the surface of this fascinating subject here, but let us look a little more closely nevertheless.

According to the massive modularity hypothesis, the human mind, or indeed any animal mind, is an ensemble of specialised modules, each of which has evolved to solve a specific problem faced by our ancestors, human or otherwise. Each module deals with a specific application, processing only information relevant to that application. It is autonomous, fast, and unconscious. Speed is of course of the essence in many situations, where efficiency of reaction, if not survival itself, may be at stake. Note in particular the idea that in many cases

[2]Naturally, this is a manner of speaking that could be misleading. Mother nature made no calculation here. These developments came about by chance in various contexts where they turned out to be favourable for the genes that predisposed the given animal to express them; favourable in the sense that those genes could then pass into the next generation and hence proliferate.

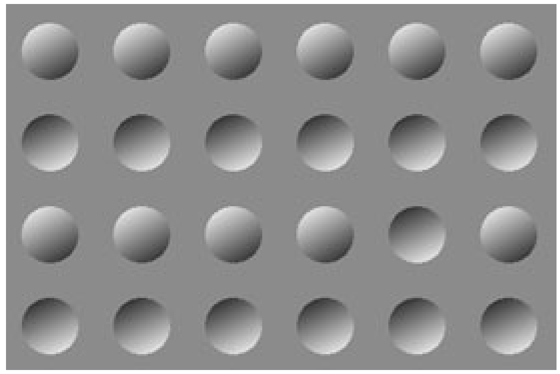

Figure 5: What do you see? The top row would seem to protrude, but this is because your mind has made an assumption about the light source—without consulting you

we are not even conscious of what our modules are doing. This is one of the great discoveries of cognitive psychology: we are actually conscious of almost nothing our brains are engaged in. As mentioned above, it is generally considered today that we—our selves, that is—are just a construction by our brain, in response to introspection, largely for the purposes of communication, and are barely involved in most cognitive activity, despite the impression our own minds may give us.

So we have modules that make inferences in many and varied domains, each taking advantage of some regularity in nature, i.e., something that always presents itself in a certain way, or something that always happens when something else has just happened. And all this occurs without our even realising it. Look at Fig. 5, which shows what appears to be a slab of material with rows of bumps and hollows on it. Most people see the top row as a row of bumps and the second as a row of hollows. That's because a module in our mind has assumed, without consulting us, that the slab is lit from above. But imagine for a moment—it requires an effort—that the slab is lit from below. Now the top row becomes a row of hollows and the second a row of bumps. In fact, we can switch from one perception to the other, with sufficient effort. Naturally, all our ancestors, human or otherwise, would have been used to things being lit from above, by the sun—a regularity in our world.

So here is an important point. A module makes an inference, i.e., it produces new information on the basis of old, but it doesn't require, or provide us with, any reason to justify its inference. It is in this sense that it operates automatically. It reasons without conscious reasoning. This is illustrated again in the famous Adelson checkerboard shown in Fig. 6. We have no *conscious* reasons to see square A as darker than square B, but we do see things like that. The two squares are identical in the image—the grey scale is the same in each (see Fig. 7)—but our brain has intervened because of its knowledge of shadows.

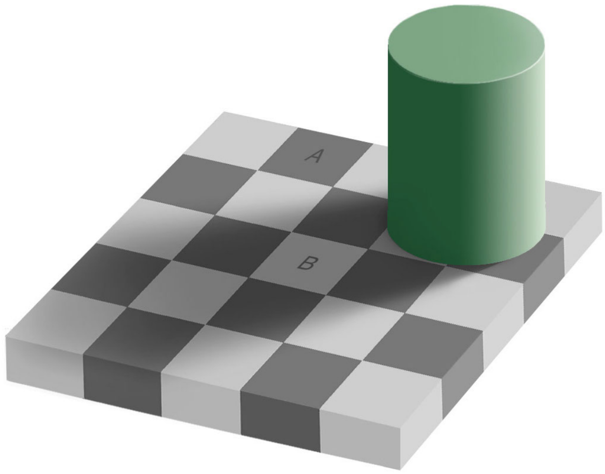

Figure 6: It may be hard to believe it, but although A is obviously a dark square and B a white one, the grey scale in the two squares is the same. Take a look at Fig. 7 to see the demonstration. Figure from [9]. Credit: Edward H. Adelson, vectorized by Pbroks13, CC BY-SA 4.0

Another experiment, or rather, type of experiment, illustrates beautifully how our brain is the final judge of the information it receives from our eyes, and quite outside any conscious effort on our part. These experiments exploit the gaze-contingency paradigm [10]: while someone is viewing a computer screen, it is possible to inform the computer in real time about the exact location of the person's gaze, and then do things to the rest of the display that the person will never be aware of. These eye-tracking techniques are widely used in psychology to study the way we view the world. For example, someone reads a short story on the screen and remains completely unaware of the fact that almost the whole of the text is scrambled at any given time. Someone standing behind the reader will find it astonishing that they manage to extract any sense from it at all.

So what has this to do with modules? In fact, a module constructs our internal image of what is visualised. In this example, the internal image bears little resemblance to what is actually there. This is because the module in question 'assumes' that what we were just looking at somewhere else a moment ago will still be there, a natural assumption in the ordinary world of nature. It exploits a regularity in the world: things don't change that quickly.

Another fascinating example of a module, which shows that some modules may have evolved a very long time ago, or perhaps have evolved several times in different species, concerns the way living creatures guide themselves through narrow spaces without touching the boundaries of those spaces. When we pass through a tunnel, the walls appear to recede at a certain speed. If the right-hand wall is closer than the left-hand wall, it will appear to recede more quickly. This experience, known as optic flow, can tell us which wall is nearer, and if we

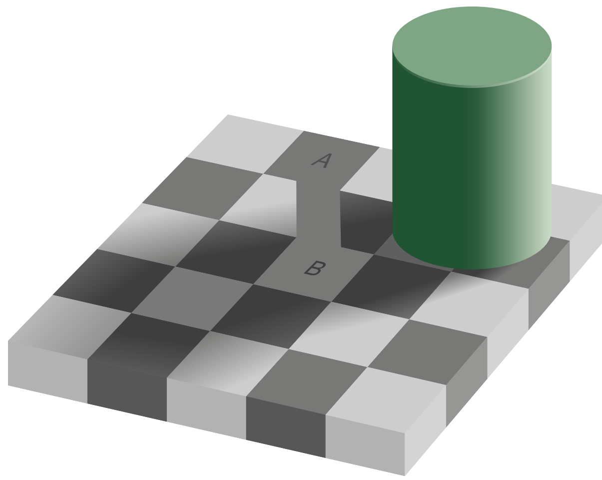

Figure 7: The grey scale is the same in squares A and B. Figure from [9]. Credit: Edward H. Adelson, vectorized by Pbroks13, CC BY-SA 4.0

wish to pass through the middle of the tunnel, all we need to do is to adjust our trajectory so that the optic flow is the same on each side. This is well described by Sloman and Fernbach in their excellent book *The Knowledge Illusion* [11, 12]:

> [One] place you experience optic flow is on the highway. The lines on the road painted by the Department of Transportation are there to keep you on the straight and narrow. As long as the flow of lines on one side of you appears to be moving by you at the same rate as the flow on the other side, you'll stay in your lane. We know this from experiments done in driving simulators. If you put someone in a simulator with a computer display and make the lines go faster on one side than the other, the person will drift toward the side with the slower lines.

Imagine a GOFAI robot doing all the necessary calculations to fly between two posts without touching them. That's okay, of course, if you can calculate fast enough. But optic flow is so much easier, provided you have some system for converting sensorial estimates of optic flow directly into motor effects.

What's fascinating is that bees use this system, too [13]. One can arrange for the bees in a hive to pass through a tunnel on their way to the meadows, then project a moving image of the wall on one side. If the right-hand wall appears to be receding more quickly, the bees no longer pass through the center of the tunnel, but more to the left. Perhaps this module evolved in a common ancestor of bees and humans, or perhaps it evolved separately several times. It's such a useful trick, that wouldn't be so surprising.

Sloman and Fernbach again [11, p. 100]:

> [All the relevant] studies show that people (and bugs) are not old-fashioned model builders that engage in loads of computation punctuated by action. Instead, people use facts about the world—like the optics of surfaces—to simplify what they do.

Modules evolve or develop when there is a regularity in the world that can be exploited to draw inferences, and when it is adaptive to do so. These inferences are not the result of reasoning in the sense we normally intend it, as you can see from the examples. They are direct consequences of one part of the brain on others. The brain applies no theory. It doesn't require any accordance of conscious thought by the self, nor indeed any involvement of the self at all.

Another example given by Sloman and Fernbach concerns the way a baseball player catches a ball [11, p. 97]. Imagine that the ball has been hit in your direction. How do you work out where you need to be to catch it? Imagine the GOFAI robot in this situation, with Newton's laws of motion programmed into it! As Sloman and Fernbach put it:

> [The robot] would sit and think about it for a while—hopefully not too long—and then move to the correct position (if it got it right).

The remark in brackets might, for example, refer to the robot's not having wind resistance programmed into it. Fortunately, the baseball player doesn't need to follow that route. Here are Sloman and Fernbach again:

> It turns out that there's an easier way to catch a ball that hardly involves thinking at all. Instead of calculating trajectories, there's a trick that will take you to where the ball is going to land. If a fly ball is coming in your direction, a natural thing to do is stare at the ball as it rises into the air, raising your head to elevate your gaze as the ball comes toward you. The direction of your gaze specifies an angle relative to the ground. Here's the trick: to end up where the ball is going to land, all you have to do is move forward or backward so that this angle is always increasing at a constant rate. In order to keep your eye on the ball after it's hit, you'll have to continuously tilt your head (or your eyeballs) upward to track the ball's motion. What may surprise you is that you'll have to continue to lift your gaze even after the ball starts to descend. If you watch an outfielder running to catch a ball, you'll see him adjust his body's direction and speed so that his gaze is always moving upward at the same constant rate. These adjustments lead him to the right spot to intercept the ball. Then all he has to do is catch it.

This is something that can be observed in the field. But with hindsight, we can also prove mathematically that it will work.

To sum up then, sophisticated tasks are not accomplished by first carrying out exhaustive planning and calculation. Our actions are often easy for us, not because the tasks in question are intrinsically simple, but because there exist simple solutions to otherwise complicated problems. High level tasks are carried out by combining simpler skills which are in turn hierarchies of even simpler skills. And we note that there is no frame problem. Each species evolves precisely the modules it needs to solve the problems encountered by its ancestors, and the relevant module always works on precisely the information it needs to take into account, ignoring the information it doesn't need to take into account. Modules evolve or develop when there is a regularity to exploit in inference and when it is adaptive to do so, i.e., when it gives the animal an advantage for its survival and reproduction. It is in this sense that evolution is opportunistic, and it is also in this sense that our modular mind is opportunistic.

Two Kinds of Reasoning

The modules just discussed confirm that it is worth considering any information processing as a form of reasoning, even when no consciousness is involved, since we can view modules as drawing inferences and applying knowledge about the regularities of the world to do so. But, of course, there is also something we call conscious reasoning. This is something we can write about or present on the slides for a lecture, or again merely explain to a colleague. In his fascinating book *Thinking, Fast and Slow*, Daniel Kahneman makes just such a distinction about thinking in general. Fast thinking is done by the kind of modules we have just described: it is rapid, automatic, unconscious, and effortless. Slow thinking on the other hand is slow, controlled, linear like the language used to formulate it, and requires a considerable effort.

Fast thinking gives rise to what are known as *cognitive biases*, and Kahneman's book is full of them. He and his colleague Amos Tversky spent much of their lives studying these, particularly in the context of economics, which is why Kahneman, although a cognitive psychologist, was eventually awarded the Nobel Prize for Economics. This psychological approach to how investors *really* invest and how financial managers *really* make their decisions has given rise to a new branch of study called *behavioural economics*, in contrast to the standard approach to economics which assumes that investors and financial managers are purely rational in some sense.

More will be said about the cognitive biases in a moment. But first let us turn to a form of reasoning which undoubtedly does distinguish us from other animals, if only because it involves language.

Moral Reasoning

We noted earlier how the Wason logic test revealed something important about our ability to be logical. It turned out that there was a close link with the issue of morality. So let us return to the question of ethics in the company of the American evolutionary psychologist Jonathan Haidt. His book *The Righteous Mind* is a great introduction to the subject [15]. Note in particular the subtitle *Why Good People Are Divided by Politics and Religion*.

Haidt sets out five, or perhaps six, evolutionary foundations of morality. This is not the place to describe them all, but let us pick out just one that seems particularly relevant here. That is the fairness/reciprocity foundation. Each foundation evolved to meet an adaptive challenge, in this case the challenge of reaping the rewards of bilateral partnerships, i.e., benefiting from cooperation. And each foundation coevolved with certain emotions, presumably unique to humans. In this case, anger against the cheat, gratitude toward those who cooperate, and guilt when we ourselves have cheated. In Haidt's view, morality is for a large part an evolved solution to the problem of the free rider. And of course, this involves reasoning about the rules, which ties in well with what was revealed by the Wason tests: our logic is sharp when it comes to holding others to the rules.

We shall come back to Haidt in a moment, but let us first mention two French authors, Hugo Mercier and Dan Sperber, and their remarkable, even groundbreaking book *The Enigma of Reason. A New Theory of Human Understanding*. It puts forward a modular approach to reason which is perfectly relevant to the subject of this article, but would take too long to expose here. However, the idea that there is something enigmatic about human reason will be further discussed below.

But first we note that Mercier and Sperber fully agree with Haidt: reasons are social constructions mainly for social consumption. Indeed, for all these authors, moral reasoning has a strategic function. We produce reasons in order to:

- justify our thoughts and actions to others,
- produce arguments to convince others to think and act like ourselves,
- evaluate the reasons that others produce to justify themselves.

The key here is of course always cooperation. We need to be trusted and we need to know whom we can trust.

Returning to Haidt, we find that he adds something else, which is particularly interesting for the present discussion. According to his assessment, *intuitions come first* and strategic reasoning second. The basic idea is that all our moral reasons originate ultimately from intuition. But what is intuition? It is an idea of some kind that we suddenly become aware of. Something that springs therefore from the unconscious activity of our brain. And that is interesting because, if Haidt is right, even this moral reasoning that distinguishes us so clearly from other animals arises ultimately from the unconscious. He gives several examples of experiments carried out by cognitive psychologists to illustrate this. Here are three:

- In a simple experiment to show how familiarity can influence our preferences [17], people are asked to rate Japanese pictograms on a scale of one to ten—as the inventors of Tripadvisor will tell you, people are ready to rate anything on a scale of one to ten. However, while they are reading the instructions on the screen, some of the pictograms are flashed up for a very short time, too short for the individual to be aware of them. Higher notes tend to be attributed to those pictograms shown subliminally. From an evolutionary point of view, it's easy to understand that it would be good strategy to favour things we know.

- Another factor influencing our moral judgement is a bad smell in the neighbourhood [18]. People are asked to judge morally dubious situations when a bad smell has been introduced in their vicinity. They don't realise that the smell has been introduced deliberately. Their judgements become harsher. Likewise if they have been asked to carefully wash their hands before beginning [19].

- And the lengths of sentences given by trained judges who have been asked to attribute punishments to criminals in case studies can be influenced by presenting them in various ways with larger or smaller numbers before they begin, once again in such a way that they are unaware of any connection between the presented numbers and the aim of the experiment. A detailed overview of this important discovery can be found in [20]. These effects are known to cognitive psychologists as priming or anchoring.

The surprising conclusion here is thus that, even though our moral reasoning is expressed in language and hence undoubtedly distinguishes us from other animals, we—our selves, that is—are much less in control of it than we imagine, because it basically surges up from the unconscious activity of our brain in the form of intuitions, or at least is heavily influenced by such. It is interesting to quote the French moralist and author of maxims and memoirs François VI, Duc de La Rochefoucauld, Prince de Marcillac (1613–1680):

> Everyone complains about his memory, and no one complains about his judgment.

So much was thus known long ago. What is new here? It is partly that we have experiments demonstrating the prevalence of the problem, and partly that, by adopting an evolutionary approach, we get a handle on understanding *why* we are like this. Or rather, *why we are not really like we imagine ourselves to be*. Basically, it is because the self is just a representation of the human animal produced by the brain for the purposes of its interactions with other selves. And there is no reason to imagine that this will take into account, or represent, the totality of the huge amount of the brain's other activity, such as why it is *really* doing what it is doing.

According to Mercier and Sperber, all our mental processes, both affective and cognitive, are now considered to be largely or even totally unconscious. Moreover, we have little or no introspective access to our own mental processes and our verbal reports of these process are often pure fabrication. Indeed, we are systematically mistaken in assuming that we have direct introspective knowledge of our mental states and the processes that produce them.

So, returning to Haidt, intuitions come first and, under normal circumstances, cause us to engage in socially strategic reasoning. In his book [15], Haidt discusses at some length the nature of this strategic operation. Here are some conclusions:

- Moral thinking is more like a politician looking for votes than a scientist looking for truth.
- What we call conscious reasoning, i.e., the linguistic formulation of our thoughts, works like a press secretary that automatically justifies any position taken by the president.
- In moral and political matters, we are often groupist rather than egoist. We deploy our reasoning skills to support our team and demonstrate our commitment to our team.
- Reasoning can lead us to almost any conclusion we wish to reach. When we hear something we agree with, we ask: Can I believe it? And it's an easy matter to find a reason to say "Yes!" When we hear something we don't agree with, we ask: Must I believe it? And it's an easy matter to find a reason to say "No!"

Here we may quote Benjamin Franklin, scientist, inventor, and one of the founding fathers of the American constitution, although he never became president:

> So convenient a thing to be a reasonable creature, since it enables one to find or make a reason for every thing one has a mind to do.

So, when language evolved, why did we develop an inner lawyer, and not an inner judge or an inner scientist? Would it not have been more adaptive for our ancestors to find out the truth about who did what and why, rather than just using all their wits to find evidence for what they wanted to believe? The answer is straightforward from the evolutionary point of view: what was most important for the survival of our ancestors was not truth, but reputation. Would we say that was rational? Well, in a sense, it was. In terms of survival and reproduction, it was essential to be well accepted, and indeed trusted, in one's own group.

Cognitive Bias and the Paradox of Reason

There is a lot of discussion of cognitive biases these days. The book by Kahneman [14] is a good introduction. Here is a definition from a website [21]:

> A cognitive bias is a systematic error in thinking that occurs when people are processing and interpreting information in the world around them and affects the decisions and judgments that they make.

The same website goes on to say:

> The human brain is powerful but subject to limitations. Cognitive biases are often a result of your brain's attempt to simplify information processing. Biases often work as rules of thumb that help you make sense of the world and reach decisions with relative speed.

Although self-contradictory—a systematic error helping us to make sense of the world?—this website certainly captures something of the flavour of these psychological effects!

Some consider cognitive biases to be a kind of oddity because they don't seem particularly rational, and this has given rise to what is known as the paradox of reason. For we may wonder how humans have managed to create such a huge body of reliable knowledge, to approximate so well to the truth about the world, as we can see in our mastery of technology. A solution to this enigma, also referred to in the title of the book by Mercier and Sperber [16], will be outlined below. But it should be borne in mind that all these biases, and dozens of them have been identified by now, are very likely universal, i.e., common to all humans, and this suggests that they have a largely genetic base; which in turn suggests that they serve some useful purpose in our quest to survive long enough to reproduce. Unless of course something has changed so radically in our modern world that they now do us a disservice.

Let us therefore take two examples to illustrate these things. The first is the *availability bias*:

> The tendency, on a given question, to privilege information that is immediately available to us.

For someone living in a small group of hunter–gatherers, this is likely to be a good strategy, since nearly all the information available concerns one's close entourage and the local environment. But the situation is very different today. Media coverage provides us with huge amounts of information, much of which barely concerns us in our everyday lives. And in fact it determines our assessments of frequency and risk. One of the reasons why people are much more afraid of travelling by air than travelling by car is that plane crashes get such a lot of coverage. What would happen if every day the TV news began by showing all the car accidents in the country? Would people drive in the same way? That is a classic example, but there is more to say and Pinker puts it perfectly [1]:

> The press is an availability machine. It serves up anecdotes which feed our impression of what's common in a way that is guaranteed to mislead. Since news is what happens, not what doesn't happen, the denominator in the fraction corresponding to the true probability of an event—all the opportunities for the event to occur, including those in which it doesn't—is invisible, leaving us in the dark about how prevalent something really is.
>
> The distortions, moreover, are not haphazard, but misdirect us toward the morbid. Things that happen suddenly are usually bad—a war, a shooting, a famine, a financial collapse—but good things may consist of nothing happening, like a boring country at peace or a forgettable region that is healthy and well fed. And when progress takes place, it isn't built in a day; it creeps up a few percentage points a year, transforming the world by stealth. As the economist Max Roser points out [22], news sites could have run the headline *137 000*

> *People Escaped Extreme Poverty Yesterday* every day for the past twenty-five years. But they never ran the headline, because there was never a Thursday in October in which it suddenly happened. So one of the greatest developments in human history—a billion and a quarter people escaping from squalor—has gone unnoticed.

Pinker goes on to highlight a very real risk for our democracies:

> Availability-driven ignorance can be corrosive. A looping mental newsreel of catastrophes and failures can breed cynicism about the abillity of science, liberal democracy, and institutions of global cooperation to improve the human condition. The result can be a paralyzing fatalism or a reckless radicalism: a call to smash the machine, drain the swamp, or empower a demagogue who promises "I alone can fix it."

So much for the availability heuristic then. In the modern world, with its abundance of often irrelevant information, this bias can easily lead us astray.

Perhaps the best known cognitive bias is the *confirmation bias*. It consists in:

- favouring information that confirms our preconceived ideas,
- according less importance to information that seems to contradict them.

Presumably any honest person will recognise themselves here. It is this bias that immediately raises the question of how we manage to produce any reliable knowledge—the paradox of reason. And since we were promoting evolution as a key explanatory tool in understanding human cognitive activity, what could possibly be the evolutionary advantage of the confirmation bias?

To begin with, according to Mercier and Sperber, we are more groupist than egoistical and the confirmation bias is actually a partisan bias. This is clear to all in political debate, and can be observed throughout human activity. But in actual fact it can provide an effective way to share the cognitive workload when a group of people have a problem to solve. For it turns out that we are lazy when it comes to evaluating our own reasons and justifications, but extremely effective at evaluating those of others. Put another way, no one is good at spotting their own mistakes, but we're very, very good at spotting other people's.

To understand how this could be useful, we must ask what is the appropriate context for reasoning. From what we have already said about the social role of reasoning, it seems plausible that it might be a discussion in which people exchange arguments and justifications and where there is a common interest in finding the truth, or at least finding the best compromise. We only have to reflect for a moment on the myriad examples of social cooperation to see that such discussions do occur, even though we may focus on situations where political debate becomes non-debate through overzealous partisanship (and by focusing on those situations, bring to bear our availability biais).

In effect, reason is much more efficient at evaluating good arguments than producing them. It is easy to show that the problem-solving abilities of a group of people are far superior in general to those of any individual. For example, groups are much better than individuals at solving Wason tasks. When people are presented with the correct explanation for the first Wason task described above, they usually recognise and accept it immediately. We return to the crucial point that the mind did not evolve in a context of individuals working alone to solve problems. We are designed to cooperate, and thinking on one's own is not a great way to get good solutions. Let me quote Keynes again:

> It is astonishing what foolish things one can temporarily believe if one thinks too long alone.[3]

One might add, if one thinks too long in a small group of people with the same ideas.

So when do we change our minds on moral issues? The answer is that, when we interact with others, reasoned persuasion becomes possible. Note then the gulf between an individual's reasoning and our collective reasoning, corrected for biases like those discussed above. And here is the solution to the paradox of reason: the best reasoning comes when we reason together, a theme I shall develop in a moment. But note first the importance of language in what we generally conceive of as reasoning, and the importance therefore of consciousness. This brings us to an age-old question: did language appear as an enhancement of communication or as an enhancement of thinking? The answer must certainly be both.

Even alone, by expressing an argument in words, we can view it more objectively by anticipating what issues might be raised by someone else. This is what happens when we prepare the slides for a lecture or write a book or a paper. And of course it is essential that the arguments prepared in this way should actually be presented to audiences and readers for their criticism. This is an important part of the scientific method, or rather, of the way scientists go about their work. Science is a collective undertaking. In fact, it is the collective undertaking *par excellence*, perhaps the greatest collective achievement of humankind. Scientists publish in international peer-reviewed journals so that their work can be criticised and corrected if necessary, and they present their work at international conferences for the same purpose.

Science implements a process of self-correction involving a whole community of people whose overall commitment is to know the reality of the world. This fits perfectly with the idea above that the appropriate context for reasoning is a discussion in which people exchange arguments and justifications and where there is a common interest in finding the truth, in this case, the truth about the real world, about the way things really are in the world as far as we can ascertain that. The result is an ensemble of reliable knowledge, even extremely reliable knowledge. There is no paradox here. This knowledge is the result of collective reasoning.

It is interesting to consider a rather unconventional view of scientific discovery. Most people associate scientific discovery with a few household names,

[3] And he concluded here "particularly in economics".

considered to be 'geniuses', like Galileo, Newton, and Einstein. This is not just a simplification, but a distortion of what really happens. These people never worked alone. Every one of them was surrounded by books and papers written by others and by colleagues, or enemies, trying to do the same thing. In the early days, of course, there were not so many scientists around, but even in Galileo's day, there were plenty of others interested in similar questions. The names we remember were just the lucky ones who stumbled on the right, or at any rate the best solutions. And if science has been moving forward with ever increasing speed as time goes by, it is certainly not because there are more geniuses at work today. It is just that there are more scientists at work.

The Knowledge Illusion

I will end this discussion by coming back to the book by Fernbach and Sloman [11]. The illusion they are referring to in the title is not some kind of illusion regarding the knowledge produced by science. Quite the contrary. The authors are themselves scientists and promote the scientific effort in their book, among other things. They are commenting on the fact that the knowledge of any individual is actually extremely limited, even though we tend to imagine otherwise. Any academic can attest to this. We are supposed to be experts in our very narrow domain, but very often, when we get a question from a student or colleague, we hesitate for a moment and promise to get back to them! We then rush off to try to remember what it is we are supposed to know about the question that has been raised. If necessary, we turn to our notes, then our library—academics always keep hundreds of specialist books for this purpose—and if that doesn't help, we start sending emails to colleagues! But in the end, because we *are* experts in our field, either we do get the answer, or we realise that a new problem has been identified.

The reality is that we are largely unaware of how little we know, because we confuse the knowledge in our heads with the knowledge we have access to. Being an expert means knowing where to get answers in our field, not having the answers there ready to use. In this connection, Fernbach and Sloman also refer to what they call the *illusion of explanatory depth*. To illustrate this, they describe experiments carried out by the American cognitive scientists Frank Keil and Leon Rozenblit [23]. In a first step, volunteers were asked to evaluate their understanding of well known things like zips, toilet flushes, sewing machines, and many other everyday items. In a second step, they were asked to describe in as much detail as possible how each item worked, including all the steps involved in its operation. For most candidates, this step never lasted very long! When asked in stage three to re-evaluate their understanding of the various items, the values fell sharply.

In a similar exercise, experimental psychologist Rebecca Lawson [24] asked candidates to complete a very cursory diagram of a bicycle (see Fig. 8) with details of the chain and pedalling mechanism. The results were amusing (see Fig. 9). And yet, if the bicycle is in front of us, we can of course explain how it

Figure 8: Diagram used by Lawson in the cycology experiment

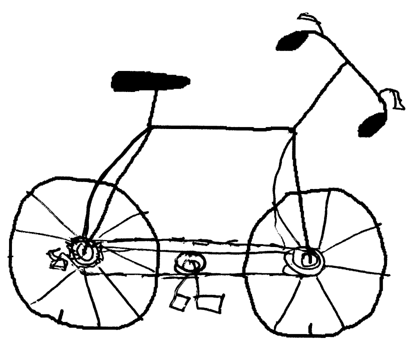

Figure 9: Example of a diagram produced in the cycology experiment

works, i.e., explain the causal chain of events between pressing on a pedal and the wheels turning. But for most of us, these everyday items have to actually be there. In a sense then, we also use the outside world as an extension of our memory and our understanding of things. The brain is efficient: it doesn't store any more than necessary.

Sloman and Fernbach have carried out similar experiments, but in a very different domain. They are concerned about societal issues like the high level of polarisation in politics, particularly in the United States, but also the often anti-scientific mood encountered among the general public. In both cases, we have a tendency to adopt very strong positions generally based on very little that we would be able to articulate. Here is a quote often reputed to be from Winston Churchill:

> The best argument against democracy is a five-minute conversation with the average voter.

What do we actually know about the consequences, the detailed consequences, the chains of cause and effect, of public policies we so harshly criticise or so generously support? This is important because each of us decides, at least a little bit. And this problem comes to the fore in the debate over representative democracy and direct democracy. Are we really up to answering the questions posed in a referendum?

To show that the illusion of explanatory depth enables people to hold much stronger positions than they can support, Sloman and Fernbach set up experiments very like those by Rosenblit and Keil. They asked candidates to rate their understanding of public policy issues that were hot subjects at the time in the United States. In stage two, they asked them to explain the consequences of those policies in as much detail as possible, and in particular to spell out the chains of cause and effect. Then in the third stage, the candidates were asked to reappraise their understanding of each policy. In the words of the authors:

> As in most experiments of this kind, participants were pretty bad at generating explanations. With very few exceptions, they simply had very little to say when we asked them to explain how a policy worked. They didn't know the mechanics of policies in a way that they could articulate. And consistent with their inability to explain, they rated their understanding the second time lower than their understanding the first time. They showed an illusion of explanatory depth. Their attempt to explain the issue revealed to them that they didn't understand it as well as they thought they had. We conclude that, just as people overrate their own understanding of toilets and can-openers, they also overrate their understanding of political policies.

However, what they really wanted to know in this experiment was not whether people suffered from an illusion. What they wanted to know was whether the attempt to explain would make them less extreme in their position on the issue. They therefore also asked them to rate their support or opposition on a scale from 1 to 7, where 1 indicated strongly in favour and 7 strongly opposed, while 4 meant they had no opinion one way or the other. And again, they asked them both before and after they explained what the consequences of the policy would be. Here is the authors' conclusion:

> We found that attempting to explain how a policy worked not only reduced our participants' sense of understanding, it also reduced the extremity of their position. If we consider the whole group together, the fact that people were on average less extreme means that the group as a whole was less polarised after the explanation exercise. The attempt to explain caused their positions to converge.

But there was one more experiment in the series. Sloman and Fernbach again:

> Usually when people think about and talk about policies, they are not engaged in causal explanation. Most discourse about policy is about why we believe what we do: who agrees with us, why we hold whatever value the policy addresses, what we heard about it on the news the other day. Our experiment asked people to do something difficult and unusual, to causally explain the effects of a policy. That task requires engaging the details of the policy and spelling out how the policy would interact with a complicated world.

So what would happen if, instead of asking for the detailed causal analysis, candidates were asked simply to give their reasons for opposing or supporting a policy? In the words of the authors:

> They were asked to state precisely why they felt the way they did about the policy. Instead of getting outside their own interests by thinking about the policy on its own terms, we asked them specifically to think about the policy from their own perspective. in this way, we asked them to do what people normally do when thinking about political policy. Participants answered the same questions as those in the first experiment: they rated their understanding of the issue and their position on it both before and after generating reasons.

Generating reasons rather than a causal explanation led to quite different behaviour: participants showed no decrease in their sense of understanding, nor did they moderate their positions.

The upshot of all this is simple: if you want to reduce the polarisation in a group, don't ask people the reasons why they adopt the position they do. Ask them to produce a detailed causal explanation. The beauty of causal explanation is that it takes those who have to explain the thing outside of their own belief systems—not just religious beliefs, but belief systems in general. Breaking the illusion of understanding reduces polarization because people realize that they don't really understand the consequences of a policy. The problem is that, when we are not qualified, we don't know what we don't know, and we don't even know that we don't know. The problem is not ignorance *per se*—we are all ignorant. The problem is that we are too quick to forget how ignorant we are.

Another issue raised by Sloman and Fernbach is the strong anti-science mood in our Western democracies, not just in some intellectual circles, but also among the general public. Why does it persist and even increase despite all our attempts to educate people? Once again, when it comes to big and complex issues like the use of GMOs, vaccines, antibiotics, nuclear energy, and so on, we do not know enough individually to form nuanced and informed opinions about such new technologies and scientific developments. And this is not just a question of whether we understand the science involved. It's the question of positive and negative impacts and the complexity of weighing up risks and benefits. In the end, as ordinary citizens, we have no choice but to adopt the positions of those we trust. And this raises the question: whom should we trust?

Very often, it is the people in our immediately community. And as Sloman and Fernbach point out, our beliefs are not just our beliefs. They are shared with our community and that makes them really hard to change. Beliefs are not isolated data that people can pick up and throw away at will. They are deeply connected to other beliefs, shared cultural values, and our identities. Rejecting one belief often involves rejecting a whole host of other beliefs, abandoning our community, going against those we trust and love, and questioning our own identity. In the confrontation between culture and cognition, the power of culture over cognition overwhelms any attempt at education.

So, yet another thing to bear in mind when we ask about human rationality: the importance of our sense of belonging to one group or another, whether it be a religious group, a political tendency, or anything else. As we said before, it is perfectly rational to want to be well viewed within one's community, and at the same time, for many issues that we all would like to pronounce upon today, the positions we adopt are barely rational at all when viewed in terms of the actual functioning of the world, be it the physical world, the political world, or any other aspect of our existence.

A Conclusion

The position adopted here is the one advocated by a *naturalist philosophy*. The aim is to understand humans without the addition of mystery ingredients. In this view, the key is the idea of evolution. The only way to understand who we are and why we function the way we do is to understood how we got here. We can speak of a kind of self-domestication—a coevolution of genes and cultural practices that has led our species to live in cooperative societies.

Concerning animal life in general, this provides the idea that animal minds may be composed of a hierarchy of modules which draw inferences based on regularities in the world, each generated long ago by the interactions of the animal's ancestors with their environment. Most of these modules operate unconsciously and many are likely to be shared between many species. These already operate a form of reasoning, in the sense of processing information, but our own such modules would not by their nature distinguish us from other animals.

The reasoning that would distinguish us from other animals involves language, consciousness, and our collective efforts, mainly aimed at cooperation. For we are indeed the world's champion cooperators, able to invent new activities at the drop of a hat and then work together on them. But for cooperation to develop to this level, evolution had to (mindlessly) invent morality. This was the unplanned aim of the self-domestication mentioned above, which allows the genes predisposing us to cooperation to proliferate against the threat of the genes predisposing us to profit from others without due return, those that shape the free rider in us. But even in the domain of moral reasoning, we are still heavily influenced by factors of which we are just not aware, a legacy no doubt of the long evolutionary process which brought us here.

So, have we become rational? Individually, we certainly have in a certain sense. Our cognitive biases are universal and serve purposes of which we are only just becoming aware, thanks to our evolutionary approach to understanding ourselves. On the other hand, this rationality is not the one we tend to think of, the one that passes through the careful logical expression of arguments. The latter has come about once again for the purposes of cooperation. Collectively, we certainly are capable of this kind of rationality, when we work together to eliminate as far as possible some of those biases. Science gives the best example, producing a huge body of reliable knowledge for the first time in human prehistory and history. This kind of reasoning comes about when we sit down together to share opposing views, but with a common aim, to reach the truth about something or to reach a compromise. In the case of science, this common aim is indeed to know and explain insofar as possible the way the real world we suppose to exist around us really is. That is, to move forever closer to the truth about our world.

References

[1] S. Pinker, *Rationality*, Viking, New York (2021)

[2] R. Baillargeon, Object permanence in 3 1/2 and 4 1/2 month-old infants. Developmental Psychology **23** (5), 655 (1987)

[3] E. Spelke, Initial knowledge: six suggestions, Cognition **50**, 431–445 (1994)

[4] M. Mitchell, *Artificial Intelligence. A Guide for Thinking Humans*, Farrar, Straus, and Giroux, New York (2019)

[5] R. Jackendoff, *A User's Guide to Thought and Meaning*, Oxford University Press (2015)

[6] J.A. Fodor, *The Modularity of Mind: An Essay on Faculty Psychology*, MIT Press, Cambridge, Massachusetts (1983)

[7] See, for example, L. Cosmides, J. Tooby, From evolution to behavior: Evolutionary psychology as the missing link. In J. Dupré (Ed.), *The Latest on the Best: Essays on Evolution and Optimality* (pp. 276–306), The MIT Press (1987)

[8] D. Dennett, *From Bacteria to Bach and Back. The Evolution of Minds*, W.W. Norton and Co., New York, London (2017)

[9] https://en.wikipedia.org/wiki/Checker_shadow_illusion

[10] https://en.wikipedia.org/wiki/Gaze-contingency_paradigm

[11] S. Sloman, P. Fernbach, *The Knowledge Illusion. The Myth of Individual Thought and the Power of Collective Wisdom*, Pan Books, London (2018)

[12] A.P. Duchon, W.H. Warren Jr, A visual equalization strategy for locomotor control. Of honeybees, robots, and humans, Psychological Science **13** (3), 272–278 (2002)

[13] M.V. Srinivasan, M. Lehrer, W.H. Kirchner, S.W. Zhang, Range perception through apparent image speed in freely flying honeybees, Visual Neuroscience **6** (5), 519–535 (1991)

[14] D. Kahneman, *Thinking, Fast and Slow*, Penguin Books, London (2011)

[15] J. Haidt, *The Righteous Mind. Why Good People Are Divided by Politics and Religion*, Penguin (2013)

[16] H. Mercier, D. Sperber, *The Enigma of Reason. A New Theory of Human Understanding*, Penguin (2018)

[17] R.B. Zajonc, Attitudinal effects of mere exposure, Journal of Personality and Social Psychology **9**, 1–27 (1968)

[18] S. Schnall, J. Haidt, G.L. Clore, A.H. Jordan, Disgust as embodied moral judgment, Personality and Social Psychology Bulletin **34**, 1096–1109 (2008)

[19] C.B. Zhong, B. Strejcek, N. Sivanathan, A clean self can render harsh moral judgment, Journal of Experimental Social Psychology **46**, 859–62 (2010)

[20] B. Englich, Blind or biased? Justitia's susceptibility to anchoring effects in the courtroom based on given numerical representations, Law and Policy **28**, no. 4, 497–514 (October 2006)

[21] https://www.verywellmind.com/what-is-a-cognitive-bias-2794963

[22] https://twitter.com/MaxCRoser/status/919921745464905728?s=20

[23] L. Rozenblit, F. Keil, The misunderstood limits of folk science: an illusion of explanatory depth, Cognitive Science **26** (5), 521–562

[24] R. Lawson, The science of cycology: failures to understand how everyday objects work, Memory and Cognition **34** (8), 1667–1675