2022

Algorithmic Rationality

Rohit J. Parikh
CUNY Graduate Center

Algorithmic Rationality

Rohit Parikh
City University of New York

October 22, 2022

Can we have thought without language?

"It is sometimes maintained that there can be no thought without language, but to this view I cannot assent: I hold that there can be thought, and even true and false belief, without language. But however that may be, it cannot be denied that all fairly elaborate thoughts require words.

I can know, in a sense, that I have five fingers, without knowing the word five, but I cannot know that the population of London is about eight millions unless I have acquired the language of arithmetic, nor can I have any thought at all closely corresponding to what is asserted in the sentence: the ratio of the circumference of a circle to the diameter is approximately 3.14159.

Bertrand Russell, Human Knowledge: Its Scope and Limits

Let me start with a quote from Andy Clark's book *Being There*.

> *Take for example the humble cockroach. The roach is heir to a considerable body of cockroach style common sense knowledge. At least that is how it must appear to any theorist who thinks explicit knowledge is the key to sensible looking real world behavior!*

- *The roach senses the wind disturbance caused by the motion of an attacking predator.*
- *It distinguishes wind caused by predators from normal breezes and air currents.*
- *it does not avoid contact with other roaches.*
- *When it does initiate an escape motion it does not simply run at random, instead it takes into account*
  - *its own initial orientation.*
  - *the presence of obstacles such as walls and corners.*
  - *the degree of illumination.*
  - *and the direction of the wind.*

Andy Clark in Being There

So the roach acts *as if* it is doing some thinking but surely not thinking as we understand it. We humans have a "knowledge base" which is typically expressed in language (although Ryle would disagree) but most creatures do not use language very much or at all. And I doubt that Yuja Wang when she plays Rachmaninoff is doing much thinking.

So we need to separate rationality from an intimate connection with language and find ways to measure it directly based on what it *does*.. We also need to be aware that for a creature or a person only *that much* rationality is required in its actions as is necessary for it to live its life successfully. Lionesses are pretty good at cooperating when hunting, but when they have had a full meal, they do not watch TV or look at their cell phone. They sleep.

Thus rationality will depend on the way they see the world, which is likely different from ours, and the abilities which they have and that rationality will be aimed at its actual needs. Rationality then should be defined as *making the best use of its abilities and its perceptions.*

We can also define a sort of IQ as the ratio of utility obtained from what a creature *actually does* divided by the utility obtained from what it *could ideally do*. This ratio will never be more than 1 but its advantage is that unlike our usual IQ, it does not depend on its chronological age .

Historically, IQ was a score obtained by dividing a person's mental age score, obtained by administering an intelligence test, by the person's chronological age, both expressed in terms of years and months. The resulting fraction (quotient) was multiplied by 100 to obtain the IQ score.

Definition An algorithmic rationality model (ARM) consists of

1. A probability space (W, p) where W is a space of points and p is a probability function on W satisfying the Kolomogorov axioms. Often W may be finite.
2. A utility function u from W into R
3. A finite set A of functions $a_1, ..., a_n$ from W to W. The a_i are thought to be actions which take the agent from point w to the point $a_i(w)$.
4. A finite set $P = \{p_1, .., p_m\}$ of perceptions which are maps from W to $\{0,1\}$. (The truth value of p_i at w)
5. A map *value* from $P \times W \to \{0, 1\}$

Explanation: Given $w \in W$, $u(w)$ is the utility to the agent (thus far unique) of being at state w.

Programs

1. Each a_i is a program.
2. If α, β are programs, so is $\alpha; \beta$
3. If α, β are programs and p is a perception then *if* p *then do* α *else* β *is a program*
4. if α is a program and p is a perception then *while* p *do* α is a program

Definition

Functions defined by programs.

- if a_i is a basic action then $f(a_i) = a_i$ (since a_i was given as a function)
- $f(\alpha; \beta) = f(\alpha)\, o\, f(\beta)$
- $f(\textit{if } p \textit{ then } \alpha \textit{ else } \beta) = \{(w, w') : p(w) = 1 \,\&\, w' = f(\alpha)(w) \textit{ or } p(w) = 0 \,\&\, w' = f(\beta)(w)\}$
- $f(\textit{while } p \textit{ do } \alpha)(w) = w'$ iff there is a chain $w = w_1, ..., w_n = w'$ such that $p(w_i) = 1$ for $i < n$, $f(\alpha)(w_i) = w_{i+1}$ for $i < n$ and $p(w') = 0$.

Intuitively each program α creates a function $f(\alpha)$ from W to W. Composing with u yields a function v_α from W to R^+. Let us call this function *the crude* value of α. Thus $v_\alpha(w) = u(f(w)$. Crude because the value depends on the w where you are now.

We now need to define the *net value* of α .

- The *safety value* of α is $min_{w \in W} v_\alpha(w)$.
- The *average value* of α is the expected value over W of v_α.
- The *best value* of α is $max_{w \in W} v_\alpha(w)$

Ordering of ARM models

Given two ARM models M_1 and M_2 we will say that M_2 refines M_1 if M_2 has more actions and more perceptions (or finer perceptions).

Proposition If M_2 refines M_1 then all three of c-value, a-value and b-value are higher with M_2.

Proof: The proof follows from the following easy observation.

Observation: Let Ω be some space of objects and let v be a function from Ω to $\mathcal{R}$. Given a subset $M \subseteq \Omega$ define $V(M) = max\{v(x)|x \in M\}$ Then if $M_1 \subseteq M_2$ then $V(M_1) \leq V(M_2)$
A creature which has more actions or more perceptions will have better values whether safety, average or maximum.

The program a_1 if p else a_2 has expected value 4.25

	p	$\neg p$
a_1	5	3
a_2	3	3.5

xxxxx

The program a_1 if $p \equiv q$ else a_2 has expected value 6. Knowing *both* p and q is better.

	$p \wedge q$	$p \wedge \neg q$	$\neg p \wedge q$	$\neg p \wedge \neg q$
a_1	7	3	1	5
a_2	0	6	6	1

Sen's Capability Approach

Sen argues that our evaluations and policies should focus on what people are able to do and be, on the quality of their life, and on removing obstacles in their lives so that they have more freedom to live the kind of life that, upon reflection, they have reason to value. The capability approach has been advanced in somewhat different directions by Martha Nussbaum, who has used the capability approach as the foundation for a partial theory of justice.

According to the capability approach, the ends of well-being, justice and development should be conceptualized in terms of people's capabilities to function; that is, their effective opportunities to undertake the actions and activities that they want to engage in, and be whom they want to be. These beings and doings, which Sen calls functionings, together constitute what makes a life valuable.

What is ultimately important is that people have the freedoms or valuable opportunities (capabilities) to lead the kind of lives they want to lead, to do what they want to do and be the person they want to be. Once they effectively have these substantive opportunities, they can choose those options that they value most.

Robeyns 2005

It is clear that a society can increase the capabilities of its members in some way. For instance a bus line makes more places accessible to people who do not have cars. Wheelchairs incease the mobility of people who are handicapped. Reserving some parking spaces for handicapped people can have a similar effect.

Thus it is society which brings about an increase from a particular M_1 to a finer M_2

Then why don't all creatures have maximal actions and perceptions?

Because such perceptions may have a cost. For instance a fly gets caught in a spider web because the thread in the web is too fine for the fly's vision. But since flies are far more numerous than spiders, it does not make sense for *all* flies to have finer vision so that a few flies survive.

A Problem

A society can create more capabilities for its members by various means. A swimming pool allows people to swim. A bridge allows them to cross a river. So each person's ARM M_1 increases to a bigger ARM M_2 and his utility can increase. *all to the good we say*.

But it is not so simple. If you have a gun it increases your utility but decreases mine. If you have a smart phone, perhaps you will not bother to talk to me.

I don't know if Sen or Nussbaum have thought about the capability approach in just this way.

We can now say that a creature is *c-rational* if it chooses a procedure which maximizes the safety value. (c stands for cautious).

We can say that a creature is *a-rational* if it chooses a procedure which maximizes the average value (a stands for average).

We can say that a creature is *b-rational* if it chooses a procedure which maximizes the best value.

In order to survive and prosper an animal needs to take some risks. Taking such risks will decrease the c value but increase the b-value. Consider a man who climbs a mountain because he hopes that a beautiful girl will love him and let us suppose also that this belief is justified. He could choose not to climb the mountain and marry a less beautiful girl. Or he may choose to climb the mountain and increase the best value but sharply decrease the safety value. This issue has been discussed in a paper by Tasdemir, Witzel and myself.

Generally there would be a conflict between maximizing safety and maximizing the best outcome But we can still define a notion of *pareto-rationality*. A creature is pareto rational if it uses a program α such that no other β exceedes α in all three of c-value, a-value and b-value. Thus consider the young man Ajay who climbs a mountain in hopes of marrying a beautiful girl and another one Surya who avoids the mountain and marries a less pretty girl. They are both Pareto rational.

Cockroaches come out at night but not during the day. We can now see that this behavior is rational. It increases the c-value since they are not visible during the night. They are also b-rational since they get to eat the crumbs which foolish humans have left behind. Overly cautious roaches may starve to death. (I won't say anything about whether *humans* leaving these crumbs are rational!).

The Tick

Figure: A tick - an unappreciated genius

Unfed adult American dog ticks can go 2-3 years (up to 1,053 days) without food!

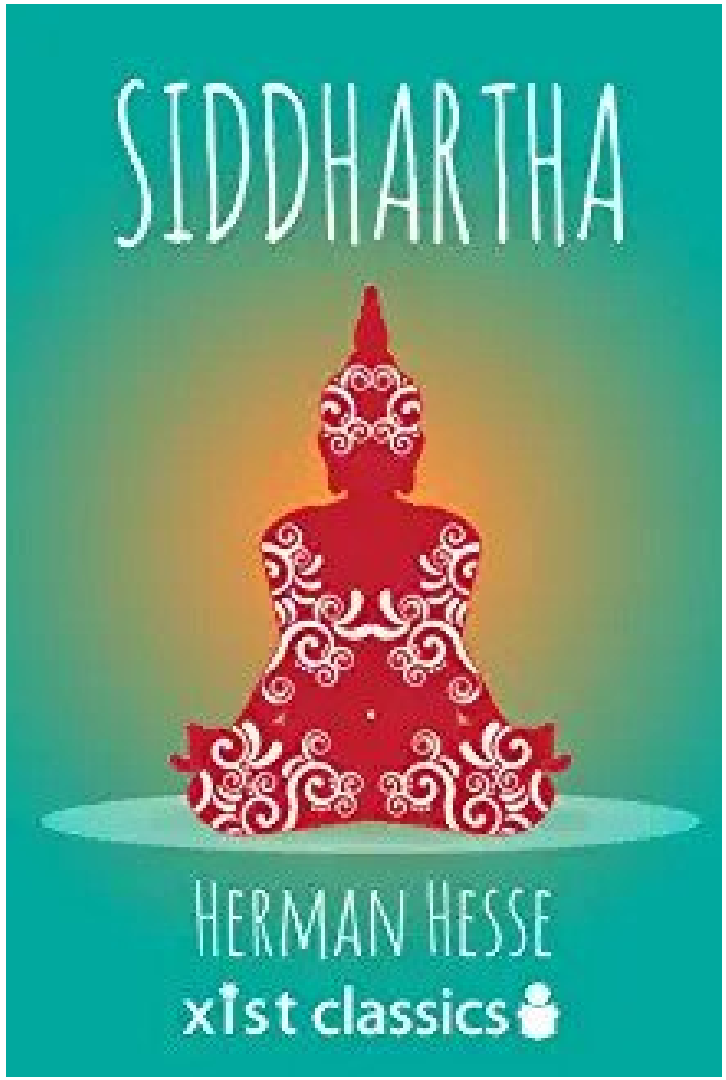

Figure: Siddhartha: I can fast and I can wait!

Once the female (tick) has copulated, she cfimbs with her full count of eight legs to the tip of a protruding branch of any shrub in order either to fall onto small mammals who run by underneath or to let herself be brushed off the branch by large ones.

The eyeless creature finds the way to its lookout with the help of a general sensitivity to light in the skin. The blind and deaf bandit becomes aware of the approach of its prey through the sense of smell. The odor of butyric acid, which is given off by the skin glands of all mammals, gives the tick the signal to leave its watch post and leap off. If it then falls onto something warm – which its fine sense of temperature will tell it – then it has reached its prey, the warm-blooded animal, and needs only use its sense of touch to find a spot as free of hair as possible in order to bore past its own head into the skin tissue of the prey. Now, the tick pumps a stream of warm blood slowly into itself.

Experiments with artificial membranes and liquids other than blood have demonstrated that the tick has no sense of taste, for, after boring through the membrane, it takes in any liquid, so long as it has the right temperature.

If, after sensing the butyric acid smell, the tick falls onto something cold, then it has missed its prey and must climb back up to its lookout post.

The tick's hearty blood meal is also its last meal, for it now has nothing more to do than fall to the ground, lay its eggs, and die.

von Uexküll, Jakob. "A foray into the worlds of animals and humans" (2010).

We will describe the activities of the tick using the following baby program.

Perceptions: *Bu (for Butyric acid), Wa (for warm), Ba (for bare) F (for full)*

Actions: *c (for climb), w (for wait), d (for drop), se (for search), su (for suck), l (for lay eggs),*

Informal program: *Climb on a grass stalk and wait until Butyric acid is detected, Then drop on to the mammal, and search for a bare spot. Suck blood until full. Then drop and lay eggs.*

More formal version: **Begin**: c;(w until B);d; if Wa then (se until Ba & Wa) else c;(su until F);d; l **End**.

It is a rather bare bones program but allows the species to prosper.

Evolution and programs

We saw that the benefit to a tick comes not from a single action but from a program. But can programs evolve? Koza 2010 says yes.

> *The goal of getting computers to automatically solve problems is central to artificial intelligence, machine learning, and the broad area of research encompassed by what Alan Turing called machine intelligence [1, 2]. As early as 1948, Turing recognized the possibility of employing the processes of natural selection and evolution to achieve machine intelligence. In his essay Intelligent Machines, Turing [1] identified three approaches for creating intelligent computer programs.*

The first approach was a logic-driven search. Turing's interest in this approach is not surprising in light of Turings own pioneering work in the 1930s on the logical foundations of computing.

The second approach for achieving machine intelligence was what Turing called a cultural search in which previously acquired knowledge is accumulated, stored in libraries, and brought to bear in solving a problemthe approach taken by subsequent work in the field of knowledge-based expert systems.

The third approach that Turing identified in 1948 for achieving machine intelligence is: the genetical or evolutionary search by which a combination of genes is looked for, the criterion being the survival value.

Thanks to Sheila Miller and Brian Skyrms for comments

References

- Dennett, Daniel Clement. *The intentional stance*. MIT press, 1989.
- Koza, John R. "Human-competitive results produced by genetic programming." *Genetic programming and evolvable machines* 11.3 (2010): 251-284.
- Lurz, Robert W. *The philosophy of animal minds*. Cambridge University Press, 2009.
- Nagel, Thomas. "What is it like to be a bat?." *The philosophical review*, 83.4 (1974): 435-450.
- Nussbaum, Martha C. "Creating capabilities: The human development approach and its implementation." Hypatia 24.3 (2009): 211-215.
- Parikh, Rohit, Cagil Tademir, and Andreas Witzel. "The power of knowledge in games. *International Game Theory Review*, 15.04 (2013): 1340030

- Rohit Parikh, Formalizing the Umwelt, talk given at the Universal logic meeting, Vichy, 2018. https://academicworks.cuny.edu/cgi/preview.cgi?article=1851&c
- Pratt, Vaughan R. "Semantical considerations on Floyd-Hoare logic." 17th Annual Symposium on Foundations of Computer Science (sfcs 1976). IEEE, 1976.
- Robeyns, I.,, 2005, The Capability Approach: A theoretical survey, *Journal of Human Development, 6(1)*: 93117.
- Sen, Amartya. "Well-being, agency and freedom: The Dewey lectures 1984." The journal of philosophy 82.4 (1985): 169-221.
- Von Uexküll, Jakob. "A stroll through the worlds of animals and men: A picture book of invisible worlds." *Semiotica* 89.4 (1992): 319-391.