2/27/2024
Xsu

Infirationality

In Economics, an individual is considered to have Superationality if that individual acts perfectly rationally with the knowledge that others also possess superationality. This can be demonstrated with the famous Prisoner's Dilemma, where two rational actors will both defect and arrive at a suboptimal outcome. In contrast, two superational actors will arrive at the optimal outcome.

In Mathematics, an introductory concept in the field of Real Analysis is Infimum vs Supremum. Simply put, the infimum ("inf" in short) of a set is the floor of that set whereas the supremum ("sup") is the ceiling.

Borrowing these concepts, a core belief of mine is what I'd like to coin as *Infirationality*. Juxtaposed against Superationality, where individuals act in a way that optimizes the systematic outcome, and Rationality, where individuals act in a way that optimizes self-interest, the Infirationality theory proposes that individuals act in a way that optimizes instinctual incentives. It may also be interpreted as an abbreviation for Infinitely-Rational.

Infirationality argues that people are fundamentally rational. There exists a set of *incentive parameters* (inputs), each weighted differently and updated dynamically, where at any given moment the *action* (output) is only performed if the expected value of the full input set results in that action. In non-math terms, all people all the time commit actions if and only if said action aligns with their incentives. The converse of this statement conveys something to the effect of people will always attempt to dissolve cognitive dissonance.

I'll illustrate with two examples, one demonstrating the converse statement and another the original.

A defeated public company CEO decides to resign after a long battle with the board. A Rationalist interpretation states that given the CEO's option set, resigning provides the highest expected value, and thus in resigning the CEO acts out of self-interest. A Superationalist interpretation states that the CEO, along with the board, all understand that the resignation lends the highest expected value outcome for the system as a whole (ie. all parties involved). An Infirationalist interpretation argues that the CEO, at the exact moment of the decision, bears extreme cognitive dissonance for a variety of reasons. Those reasons could be everything from the CEO knowing deep down that her resignation was inevitable to a temporary hormonal imbalance resulting from her lunch, which diverted blood flow from her brain to her gut, creating an implicit incentive to go with the decision that allowed her to go home and nap. Although not all of these factors are equally weighted, the cumulative expected value of the incentive parameters, at that moment in time (and could change a moment later, also known as regret),

makes resigning the rational decision. By choosing that decision, the CEO resolves the cognitive dissonance at that moment.

As I'm writing this, I am sipping on a mug of green tea. The frequency in which I choose to pick up the mug, raise it to my lips, and take a sip may appear random. However, as an Infirationalist, I believe that the action is not random at all. Instead, it is completely rational to take a sip at 3:36pm but not 3:35pm. The incentive set in this case may be: my neurological pathways that collectively represent the sense of thirst, the implicit awareness of the temperature of the liquid and its rate of decline against time (I don't want to burn my mouth but I don't want to drink cold tea either), the concentration of caffeine in my blood, the fact that I wrote the sentence "it is completely rational to take a sip at 3:36pm but not 3:35pm" at 3:35pm, and many others. I cannot identify all the incentive parameters nor can I explicitly identify the weight I have assigned to each, but subconsciously I am running this expected value algorithm constantly to decide whether to sip or not.

At this point you may think that this is a kind of "no shit" philosophy, or that I have circuitously explained a moot point. But I believe the implications of Infirationality are profound. The core belief that "people are fundamentally rational" reduces the world to a data collection and processing problem. If you can understand the exact incentive parameters and weights at every moment (important because they change second to second), you can perfectly predict how an individual will act, and at scale how every individual will act. I'll explore a few applications of this philosophy.

Trading
Early Wall Street traders were able to rake in massive profits because they had a better understanding of the incentive parameters and weights that go into the stock than the market. This was often an information arbitrage - professional traders hired people to be physically on the trading floor shouting out prices, so they would receive incentive parameters moments before the rest of the market. They were also better at calculating the weights to those parameters, as they have more experience and reps with how the market has historically moved and responded to certain news. In this case, the market is Infirational. Its incentive parameters include quarterly earnings, specific news, other correlated stocks' performances, etc; its weights are calculated through a crowdsourced process based on that stock's traders; and its actions are how the stock moves second to second. By being better at data collection and processing, traders profited when they made correct predictions on the market's output.

Habits
We can apply this philosophy to habit building. One of my goals for this year is to work out most days of the week. Having identified that action - changing, walking to the gym, working out - as the desired output, I can now collect and process data on my incentive parameters. One such parameter is the observation that I experience discomfort when I'm on the treadmill with a full stomach. Another is the observation that if the latency between the thought of "I should workout" and being physically in the gym is long, I tend to not do it. A parameter that didn't previously exist but I can establish is some sort of chemical reward to working out. So I adjusted my diet to

eat one meal a day at night such that around the afternoon I am on an empty stomach; moved to a building with a gym ten floors below my apartment; started wearing sweatpants daily to avoid having to change for working out; and started chewing nicotine gum right before going to the gym. Now I go to the gym about six times a week because it's fundamentally rational for me to do so.

Policy

Policies are a forcing function that artificially adjusts the weights of certain incentive parameters. Take the example of a speeding ticket. It may be illegal to speed on the highway, but nothing is stopping any single driver from doing so. The reason most people slow down when driving past a speedometer is that at that moment in time, when the driver recognizes the upcoming speedometer, the weight of one incentive parameter - the negative consequences of getting a speed ticket - is raised significantly. After passing, the driver likely accelerates back to his previous speed. Now imagine the driver is driving in a gated community with lots of speed bumps. In this case, it is not rational for the driver to speed since the weight on another incentive parameter - the discomfort from driving over a speed bump - is very high. Once the driver is out of the community and on regular roads, that weight is brought back to a normal level. Now people still speed through speedometers and over speed bumps, because the forcing function in this case is relatively weak and the affected incentive parameters may not be enough to overcome other stronger incentive parameters - the need to catch a flight, for example. Through this lens, I marvel at how civilized of a world we live in. When the Starbucks barista hands me my coffee, I don't even remotely consider the possibility that she may pull out a gun and shoot me in the head. That's because the laws we have in place make it fundamentally irrational for most people to do that, but that does not mean she cannot physically do it. Geopolitically, our sanctions against Russia act as a forcing function that artificially changes the incentive parameters of Vladimir Putin. However, since Putin is still waging war, it must be because it is still rational for him to do so. I speculate that we don't understand enough of Putin's incentive parameters and have been attacking it in a very unidimensional manner. If, for example, we were able to kidnap Putin's family and hold them hostage, then I'm sure it becomes a whole lot more rational for him to come to a compromise. Infirationality reduces the problem of ending the war to the problem of collecting more data on what Putin's incentive parameters are and how he weighs them.

Consciousness

The nature of consciousness itself may be highly correlated to the dimensionality of the being's incentive parameter set. We as conscious humans have so many incentive parameters that it's very hard to decipher why exactly people do certain things - and thus we wrongly convict them as "irrational". The bots in the Game of Life, on the other hand, have very clear hard-coded incentive parameters. Their actions are perfectly rational outcomes from following those incentive parameters, just like us, but we don't deem them conscious because we can understand exactly why they behave the way they do. The higher the number of incentive parameters a being has, the lower the observability we have in how they arrive at their outputs, and when we don't understand that computation process, we say it is irrational or random. A cat is "rational" most of the time - if pet, then purr; if loud noise, then run - but also can appear pretty

random when they knock things over or zooms around the house at midnight. Those behaviors aren't random, we just don't understand enough of the cats' incentive parameters much less how they weigh each one. Now take a goldfish - if poke, then swim away; if food, then eat - they seem much less random than cats. Most people will probably agree that cats are more conscious than fish, and fish are more conscious than the Game of Life bots. It can then be concluded that the number of incentive parameters is at least correlated to the level of consciousness of a being.

This is the point where we have two massive God-shaped and AI-shaped questions.

God
Following the line of reasoning, an omniscient being that is infinitely more conscious than we are must have an equally infinitely large number of incentive parameters. This means that this being will appear completely "irrational" or random to us since we will understand nothing about the computation that led to the action sets.

AI
Neural networks are Infirational where the weights are basically incentive parameters, and the biases weights (poor naming I know). Symbolic AIs are perfectly understandable, whereas Neural Net AIs are less understandable. Symbolic AIs are definitely closer to the Game of Life bots, but where do we put GPT-4 on the scale of goldfish to cats to humans?

Maybe for another time.