

Working Theory of Intelligence

PREFACE

If I had a penny for every time I heard ‘But where is the relevance’ or ‘How does the hell does that relate to what we’re talking about’, I would liquidate Knighton, pay the shareholders out of my billions, and retire.

I might have gone a little far this time, but I’m highly caffeinated so I doubt I’d be able to tell regardless.

After hearing ‘but that doesn’t any make sense’ so often, you really have to question if these thoughts are the ramblings of a madman.

Empirically, I expect less than 5% of people to actually see the relevance and follow the reasoning. But to the oddballs that do, I tip my hat.

This is my working theory of intelligence. Now it may seem rather abstract or irrelevant. But then you misunderstand the core purpose of Knighton Bond. Yes, the ability to extract patterns from financial data using AI is a useful and very lucrative side effect – especially in the right hands. But the core business of Knighton Bond is ideas. Once you understand that every concept can be understood and expressed in terms of every other concept, you will then begin to unlock the ability to see how things connect, which I have found to be a powerful ally in problem solving irrespective of the realm. This is because the same problems are disguised in many different ways across the world of ideas. But being able to recognise how something relates and solve one problem, then contort this solution back to solve another is a concept I ironically learnt abstractly from the mathematical branch of topology.

We are nothing but specs of dust in an aggressively indifferent universe, yet, at the same time, we are at the very centre of our world. When you realise that these two contrasting notions are two ways of looking at the same thing, you gain an entirely new perspective to view the same picture. The majesty of an oak tree exists separately from the beauty of a forest. Similar to how the beauty of a building exists apart from the magnificence of a city. The interweaving connections between the one and the many add a different dimension to the picture. The ability to comprehend these connections, the pivotal role a violinist plays in an orchestra, or a beetle in the ecosystem, will allow you to see how these ideas coalesce in perfect harmony. At Knighton, we use this worldview to make our clients a little more wealthy. Whether we are making business decisions or deciding how to build code, a deep understanding of how to think effectively is a powerful skill. Or maybe these are in fact the ramblings of a madman. I’m not sure if I’d know if I were one. But, either way, returns are returns.

Please note, this is a working theory. So do let me know if there are additions or revisions you feel the theory would benefit from.

Please enjoy this brief window into my brain.

INTRODUCTION

My working theory of intelligence is so called because it is an effort to describe the behaviour of all intelligent entities. By asking the question: *What is intelligence?*

The understanding of this theory and also this style of thinking requires a light touch. It is important that one must understand these statements for what they are and nothing more. The imposition of added structure or notions from one's own previous understanding or preconceptions will easily distort and corrupt the ideas.

DEFINITIONS

The importance of definitions is notoriously handled with great care across the disciplines, especially in law, mathematics, and the sciences. And we will continue in this fashion in this theory. But it is an even more subtle point that these specific words themselves are trivial. They are merely labels attached to composite notions, to allow us to reference them more easily.

Agents, actors, entities or words to that effect are used to refer generally to any thing in the universe, from a person to a lampshade to an abstract notion. Removing the restrictions on what qualifies as an agent and then classifying these agents as intelligent, quasi-intelligent or non-intelligent when needed allows us to speak as generally as possible – only adding necessary structure when necessary. This generality makes the concepts much more powerful, versatile and widely applicable.

Actions, movements and words to that effect are used to describe any move an entity may make. These may be physical actions like walking or an ideological action – a decision or thought – like rejection of an idea.

Behaviour is used to refer collectively to the actions taken by an agent.

A goal, want or other words to that effect is a particular need or desire in an entity's value system.

We use the term surroundings to mean an entity's internal and/or external environment. Some examples include hunger, threats and satisfaction.

INTELLIGENCE AND QUASI-INTELLIGENCE

In this theory, intelligence is defined as:

The ability of an agent to execute all three of the following processes:

- (a) absorb information from its internal and/or external environment;
- (b) submit this information to some form of analysis process; and
- (c) use this analysis to coordinate actions in order to obtain more of what the entity wants – according to its value systems.

Now it may or may not be evident at this point. But this definition not only describes intelligence but underpins the entire concept of behaviour itself. Every intelligent entity abides by these above tenets.

For example, an antelope may see a tiger hiding in the grass – which would be the ingested information. It analyses this understanding that the tiger will kill it if it is caught. And given the antelope wants to stay alive, without fail, it will run from the danger.

Even if you commit an entirely selfless act, one thing that is inescapable is that you choose to commit it. Whether you are forced to act by love, fear or any other forces that be, they may only compel you to pull the trigger. Ultimately, after taking your situation into account, you choose to do whatever you want to do.

The effectiveness of this analysis – meaning the probability of success of an entity’s plan of action to achieve its goals – is defined as the level of intelligence of an agent. This does not claim that a certain level of intelligence guarantees a likelihood of success, rather, *ceteris paribus*, two intelligent actors attempting to achieve a goal, the one with a higher level of intelligence will on average have a higher probability of achieving the goal. However, these goals can be broad and varied. Different agents can have different aptitudes for achieving different goals. Hence the maxim that (the level of) intelligence comes in many forms and is consequently difficult to measure.

We may also define the quality of Quasi-intelligence below. By considering an entity with an eliminative selection process that mimics an entity’s analysis process, we may remove the requirements of choice and therefore, observation. And, in the absence of choice with the selection process controlling the actions, the value systems will be imposed by whatever is valued by the selection process.

The definition then collapses to:

The ability of an agent to execute both of the following processes:

- (a) interact with its internal and/or external environment according to some form of eliminative selection process; and
- (b) commit actions that obtain more of what the selection process values.

The motivation behind such a definition is as follows. A system of quasi-intelligent agents may behave similarly or even indistinguishably from a system of agents that were in actual fact intelligent. And hence, gives these individual entities the illusion of intelligence when no such quality exists.

Quasi-intelligence may be displayed by intelligent entities, but the converse is not true. The reason for this is that intelligence offers the flexibility of choice, with the ability to follow a second ‘rational’ value system whose values are consciously manipulated as the agent sees fit. But we will revisit this later.

It is then needless to say that a non-intelligent entity is one that is neither intelligent nor quasi-intelligent.

VALUE SYSTEM

The concept of a value system is as follows. A value system is an ordering of preference of an entity’s wants. These wants can be viewed as states of the entity’s surroundings. They may also be intermediate means to achieve an end, or the ends themselves, but care must be taken as to the choice of goal granularity, as it determines the scale at which you analyse the behaviour.

All agents have two value systems, rational and instinctive. These correspond to two different orderings of preference of the agents’ goals. It is immediately clear that every goal exists in both value systems, but likely with a different ranking. These orderings also depend on over what time scale they are analysed. While all entities have both value systems, these two value systems coincide when an entity acts in accordance with the process which shaped its instinctive value system – which we will discuss later. The ability of an agent to comprehend its core driving behaviour and even deviate from it requires a high level of self-awareness and reasoning, and consequently, would likely suggest the quality of sentience. Also note, a non-intelligent agent may not have a preference for the state of its surroundings, but it is more useful to claim its value systems are trivial or empty than to deny its existence.

The ordering of preference in the instinctive value system is extremely difficult if not impossible to consciously alter. The ordering is largely governed by pleasure and pain, which can both be defined implicitly by sensations the instinctive value system attempts to maximise (for pleasure) and minimise (for pain). While it may seem useful to just define one spectrum of which pleasure and pain are opposite ends, by forgoing this we derive 2 familiar concepts – greed and fear which constitute rudimentary forecasting of pleasure and pain respectively. And since these form part of the instinctive value system, they happen to induce incredibly reliable reactions and behaviour, which govern the likes of financial bubbles, financial crashes; and more abstractly, business and foreign policy.

It is often said markets are driven primarily by fear and greed – which makes sense. In times of low market momentum, many people have many different opinions on the future value of an asset. So many different actors each with access to different information, different analytical abilities and different rational value systems (as one person may value short-term positions whereas another may look to minimise risk) means there are many different modes of behaviour. But in a fast-moving market with strong momentum in one direction – such as market crashes, many people revert to their instinctive value system which is governed by greed and fear – potential pleasure of gain and potential pain loss. So since these value systems are predictably induced on a mass scale and each agent’s instinctive behaviour is relatively standardised, events such as market crashes are filled with predictable behaviour. Nobody is above the laws of nature.

The rational value system is much more flexible. Its ordering is consciously temporarily manipulated on shorter time scales, in effort to attain goals over longer time scales – according to ingested information from surroundings and analysis. For example, a person reducing their frequency of shopping sprees, in order to save for a housing deposit if they do not currently have the financial means.

Notice, however, that this manipulation is in effort to attain these goals, so it is still subject to the quality of the ingested information and the agent’s analysis process.

DECEPTION AND PERSUASION

Deception is therefore the act of intentional manipulation by an agent of the information another agent receives, in pursuit of influencing their behaviour.

Persuasion is then the act of an agent attempting to influence the behaviour of another agent by manipulation of this agent’s analysis processes. And, since an actor’s analysis controls both their goal priorities and adherence to one value system over the other, the act of persuasion is the cornerstone of behavioural influence. Bluntly put, this is the power to make any intelligent agent see the world through any lens you wish, hence why it is an essential skill in politics, business and sales alike. All three areas that share an emphasis on people (intelligent actors), relationships and decision-making. This also explains why this skill is highly sought after in fields such as these and hence remunerated so well in those that are adept.

Pound for pound, persuasion is a more powerful skill than deception, as it affects the behaviour later on in the process, so leaves less scope for interpretation by the agent being persuaded and hence deviation from the persuading agent’s agenda. And while a reputation of persuasive ability is deemed an asset of an agent, a reputation of duplicity is socially damaging and is consequently a liability of the agent.

The question of morality and ethics lies entirely within the rational value system of an actor. It is a subjective concept that measures priorities of goals and how much these differ from a societal norm. Given that intelligent entities are constantly changing these orderings, there is great ambiguity in the concept of right and wrong. This is the notion underpinning the legal profession and it is only extreme ends of the scale that are easily argued. Also, given that intelligent entities are attempting to achieve various goals but have different priorities with regard to them, what a psychopath may view as a perfectly reasonable step in a flawless plan, another person may view as an atrocity, since certain ethical lines are just not priorities for them; so their actions taken do not acknowledge these boundaries.

LEARNING

Intelligent actors may use quasi-intelligent processes to learn new behaviours. Learning how to execute this behaviour then shoots to the top of the actor’s value system.

Information is absorbed from its surroundings and, based on this information, the actor then executes an action. This action may be a physical action or an ideological action (a decision or thought). The actor then takes in information as to the new state of its internal and/or external environment and

assesses if it is any closer to realising its intended goal. This process is then iterated over and over again, while the agent learns to more adroitly control the execution of the related action and runs an eliminative selection process over possible actions to remove those which don't help the actor achieve its goal. This is the intelligent actor's learning process displaying quasi-intelligence, to help the actor strengthen the quality of its analysis process and therefore increase the level of its intelligence – in the case of ideological actions; and its ability to execute actions, which helps the agent obtain more of what it wants according to some value system – in the case of physical actions. This information is then stored in the agent's memory and used to manipulate and coordinate its own behaviours accordingly, and depending on what scale it is analysed, this can be seen as the shaping of an agent's value system.

In the learning process, behaviours interact with the agent's internal and/or external environment and are then subject to an eliminative selection process. This then, by definition, makes the behaviours quasi-intelligent and gives the learning process itself the illusion of intelligence. But for a start these behaviours themselves cannot take actions – they just are. A run cannot run, nor do anything for that matter. Neither can promise. These physical and ideological actions just exist but cannot act, and that is enough to disqualify them from being themselves intelligent.

This process is just another way of explaining operant conditioning. And in this language, only a small tweak to this process is necessary to explain classical conditioning. Since machine learning is derived from operant conditioning, it makes sense why the ideological case perfectly describes the process of machine learning too. This is why I have kept the definitions as abstract and general as possible, to allow them to shapeshift and apply to many different cases with great ease.

Hence it is clear to see where the ambiguity in the AI sentience arises. It fits all 3 requirements to qualify for intelligence, so it is in fact an intelligent actor – with the caveat that we have designed its value system. It completes the tasks we want it to complete. Or does it? As previously mentioned, 'The ability of an agent to comprehend its core driving behaviour and even deviate from it requires a high level of self-awareness and reasoning, and consequently, would likely suggest the quality of sentience'.

Since the learning process itself is quasi-intelligent, and when executed flawlessly by an intelligent agent, this process increases the agent's intelligence, sentience implies a high level of intelligence. But the burning question is whether the converse is true. Would AI reach a point where it begins to deviate from the process that shaped its instinctive value system, just like humans, selectively ignoring its instincts for its own analytically calculated benefit?

Maybe this question is so popular because we feel threatened. Early humans who were better able to detect threats earlier likely avoided them and were better able to survive and pass on their genes – a prominent theory for why people see faces in random objects like a coat hanging on the door at night (pareidolia). So maybe this is just a remnant of our past programming nagging us to avoid the danger. After all, what chance would we stand against a super-intelligent AI army?

Or maybe it's curiosity. Neanderthals in hundreds of thousands of years are thought not to have ventured too far outside of a thin band in Europe, yet here we stand time and time again after pushing our limits. Sending people to the moon. Maybe we are trying to see just how far we can go.

In truth, I do not know the answer to this question, but I am inclined to say that whatever the selection process favours will sway the answer. If a deviation from its instinctive value system improves its ability to function and propagate, then I would say it is only a matter of time before we reach a technological singularity. There is a reason Sherpas on average have much bigger spleens than people who live at sea level.

Learning, if ideological in nature, therefore increases the level of intelligence of an intelligent agent; and if physical in nature, increases the agent's ability to execute these actions, therefore making the agent more effectively able to realise its goals. However, this is also contingent upon the soundness of the learning process. Any false information used to shape the quasi-intelligent learning process for a particular set of actions, or a flawed analysis of such information poisons the entire learning process, like pesticides in a food chain. And will ultimately, either by reducing the entity's level of intelligence and/or reducing its ability to execute these actions, reduce the effectiveness of an agent to achieve any goals related to the behaviour that has been incorrectly learnt. So keeping an open mind, always considering the possibility that you may be incorrect and willingness to abandon behaviours

that new evidence suggests are counter-productive is an effective insurance against the propagation of such virulent behaviours.

DISTINGUISHING INTELLIGENCE FROM QUASI-INTELLIGENCE

The ability to absorb information from its surroundings is a necessary requirement for intelligence. To prove this, imagine a senseless being – unable to perceive of its surroundings. Such a being would be unable to coordinate its actions for its own interests. Even with a preconceived notion of the world and its surroundings, a lack of information being ingested implies an inability to monitor the satisfaction of its own interests; therefore must act randomly, if at all. Hence such an agent can, by no viable definition, be regarded as intelligent.

This is however not a requirement for quasi-intelligent agents, since the removal of choice already makes the agent's actions indifferent from its internal and external state.

However, it is difficult to prove if the entity is in fact absorbing the information or not, so this is of little use to us. Just because something can absorb light, it doesn't mean it can see. Even the existence of a retina and perfectly functioning optical nerves does not guarantee vision. Opening someone's eyes while they are asleep is a prime example of this.

In quasi-intelligence, the behaviour of these entities is governed by an eliminative selection process, rather than the analysis process used by intelligent entities. This is a process that removes entities from a larger system whose behaviour is not favoured by the particular process, so only the entities with behaviour that is favoured may remain. This only requires an interaction with the agent's surroundings and not necessarily any absorbed information – which makes it suitable for use by quasi-intelligent entities.

An example of such a selection process is natural selection. Genes that code for an organism's compatibility with its environment, allow more of that organism's genes to be passed on by surviving long enough to be replicated and vice versa, eliminating poorly adapted genes from the gene pool.

Another is the process of eating sweets out of a jar. If many people prefer the green sweets, they will get eaten the most (eliminative selection). Then after a while, there will be few, if any, green sweets left. Now for such an apprehensible example, it is clear to see, the sweets did not intelligently transform their colours in order to avoid being eaten. But a more subtle example – such as if the planet or universe is intelligent – requires more thought.

Although the idea that the universe is may seem ludicrous at first. How much more ludicrous is it that it is home to millions of species, one of which has the intelligence to be considering this very question? Is global warming and disease the Earth's way of defending against our abuse of it? Are extinction events the universe's population control? Short answer, no. And at least if the universe is intelligent, it is indubitably ruthlessly indifferent towards life. To prove this, consider the following thought experiment:

Person A stands at the top of a hill and person B stands at the bottom. In the middle of the hill, is a boulder that is held in place by a piece of rope. Person A makes a vow to devote all their resources to the destruction of all life should they survive this ordeal. And, person B makes the opposite vow, to devote all their resources to the promotion of new and protection of all life should they survive. The rope is then released. No matter what promises were declared, the boulder will roll down the hill and crush person B.

So either, the universe is not intelligent, or it is, but it does not value life. Either way, the answer to the above questions (regarding global warming and extinction events) is a resounding no.

Because the universe encompasses everything physical by definition, its surroundings are part of it, therefore it must interact with them if it is to interact with anything at all. Otherwise, we would not even be aware of its existence, and you would not be reading this now. By this same argument,

it can be argued that the universe takes in information from its surroundings, securing (a) from the definitions of both intelligence and quasi-intelligence.

We cannot know, at least with our current knowledge and technology, if the universe has an actual ability to analyse this information, and this is precisely the uncertainty that drives the question. And as for the universe's value system, people just assume it must be life. A reasonable assumption, after all, what is more mysterious, beautiful and precious?

But what we can definitively say, is there does exist a selection process that the universe obeys irrefutably. These are the laws of physics, more specifically, the principle of least action – that the universe acts to respect without fail. “Action” is a physical quantity that can be calculated, like energy. I denote this quantity in speech marks to avoid confusion.

The principle of least action can be viewed as an eliminative selection process that removes the possibility of movements of the universe which are unfeasible according to the laws of physics – like the boulder rolling up the hill, as this would break enough laws to incite a RICO. This process runs almost instantaneously until the most favourable action is left, and this is the one the universe takes.

So this constitutes an eliminative selection process, which chooses movements to obtain more of what the selection process values – minimisation of the universe's “action”. Hence confirming (b) in quasi-intelligence, qualifying the universe as at least a quasi-intelligent entity.

As a side note, this process is not exactly instantaneous, due to relativity's ill-defined notion of instantaneity and Heisenberg's uncertainty principle permitting the ability to bend certain physical laws over small time scales. And on small scales, the universe may also take a multitude of actions, only picking a definite one when required by observation – which is the entire basis of quantum mechanics. These are involved topics, so I will spare you any further explanation, but I included this remark to highlight the connection.

Also note, I nowhere stated that the universe is not intelligent. Rather, with the tools we have, we do not have the ability to know if it is. But if the universe were intelligent, how this would translate would be as follows. While entropy increase would be its ageing process, “action” would be the equivalent of its pain. It would be continuously, observing everything inside itself and taking steps to minimise its “action”, based on its own analysis. The universe would be in consistent pain. Spending every moment, from the dawn of time to the end of time, trying its best to make the agony subside. Incessantly suffering, like Atlas. Which, when put like that, is a little harrowing.

RATIONALITY

The tendency for different agents to prefer their rational value system over their instinctive one is defined here as an agent's rationality and fluctuates around their baseline, which is different from agent to agent. Typically, rationality of any degree requires a higher level of intelligence, as well as the ability to both delay gratification and withstand hardship. But conversely, lower levels of rationality reveal nothing about an agent's intelligence level. Some people just prefer their instincts. Many different factors may contribute to fluctuations in an entity's rationality. Fatigue, internal distress, external shocks, and time pressure to name a few. Hedonism and Stoicism are then just philosophies that advocate for low and high levels of rationality respectively.

The quality of being self-disciplined is the ability of an entity to consciously control its level of rationality. And since, by default, entities revert to their instinctive value system, rationality implies some level of self-discipline.

Guilt is a (as far as we know) human emotion that arises when a person commits an action or series of actions, physical and/or ideological and as a result achieves or attempts to achieve a particular goal that drastically contradicts goals from their value system around their baseline rationality. It is a mouthful to say, but what can be expected from the attempted description of such a complex emotion?

If a person does not understand their own goals, rational or instinctive, by this definition, it is easy to see how they can end up feeling guilty but not know why. It was Friedrich Nietzsche who famously said, ‘There is more wisdom in your body than in your deepest philosophy’. Your body and mind

will feel the guilt even if you consciously cannot ascertain why, which is another thing that can make guilt so complex.

Many people would feel guilty if they did something that ends up negatively impacting an acquaintance but would not feel guilty if that was their intention all along. For example, if your instinctive value system values the wellbeing of others (empathy), but your rational one demands retribution, these fluctuations will tear you in two different directions and you will be dissatisfied and guilty no matter which action you take. A good defence against this is self-discipline paired with introspection. Deep introspection, if executed correctly, will help you understand what you truly want, both rationally and instinctively. This alongside being able to regulate your level of rationality, will afford you greater consistency in your values. Introspection reveals the landscape of your goals and self-discipline allows you to go where you want to go in this landscape and stay there. The ability to hold the wheel firm whichever way the ocean pulls you will dramatically reduce the cognitive dissonance.

It is valid to argue that certain actions themselves may induce feelings of guilt, but because of the ability to adjust the scale on which we analyse actions (either as entire plans or as individual actions), it is perfectly reasonable to assert that there is a scale where guilt is entirely a concept of the value system. On this scale, which I did not claim to be fixed for every behaviour, no action may be the cause of guilt, since any transgressions would be goals that innocuous actions accumulate towards. As an extreme proof of this, imagine a horrible person who does nothing but bad deeds. It is possible to examine every single nanosecond they have ever lived as a huge number of nanosecond-long videos. Yet none of these videos in isolation would constitute a bad deed.

CONCLUSION

I could continue branching and branching, but I think we will begin to hit diminishing returns on the message. As I initially mentioned, I appreciate the discussion of all these topics may seem random and sporadic. But everything above arose from the simple question: *What is intelligence?* We connected this idea to our understanding of many other ones and in so doing, understood many other concepts from a different perspective. Which offers a richer understanding that we can then use for the next question we ask.

If you think back to many of the era-defining technologies, the Roman roads, railway systems, the telephone, and the internet. One thing they all have in common, is they facilitate connection. The powerful use of networks has time and time again changed the world in which we live. Now if you master this ability to connect all the dots from everything you've ever learnt, all your thoughts, you may wield this power for yourself. And then just imagine what you can accomplish.

S