Leibniz, Digital Philosophy and UAI (manuscript)

Xi Li

Center for Logic, Language and Cognition

Department of Philosophy, Peking University

Beijing, 100871, China

xli@pku.edu.cn

23rd September 2013

Contents

§1 An Overview of Leibniz's Philosophy

§1.1 Ontology

> *"It follows from the* ***supreme perfection*** *of God, that in creating the universe He has chosen the best possible plan, in which there is*
>
> - *the greatest variety together with the greatest order;*
> - *the best arranged time and space;*
> - *the maximum effect produced by the simplest means;*
> - *the highest levels of power, knowledge, happiness and goodness in the creatures that the universe could allow.*
>
> *For since all the possibles in the understanding of God laid claim to existence in proportion to their perfections, the actual world, as the resultant of all these claims, must be the most perfect possible. And without this it would not be possible to give a reason why things have turned out so rather than otherwise."*
>
> — *Leibniz*

Leibniz based his philosophy upon **two logical premises**, the law of contradiction and the law of sufficient reason. The former is sufficient to demonstrate all mathematical principles. But in order to proceed from mathematics to natural philosophy, the latter is required. The law of contradiction states that all analytic propositions are true. The law of sufficient reason states that all true propositions are analytic. He also inferred from the Principle of Sufficient Reason that there are not in nature two real, absolute beings, indiscernible from each other.

Because of the principle of sufficient reason, **everything happens mechanically** in nature, that is, according to certain mathematical laws prescribed by God. "As God calculates, so the world is made.", "Unless physical things can be explained by mechanical laws, God cannot, even if he chooses, reveal and explain nature to us."

God will have perfect **power**, **knowledge**, and **will**; that is to say, God will have **omnipotence**, **omniscience**, and **sovereign goodness**.

God can bring a real world into existence merely by decreeing it. When creating the worlds, **the simplicity of means should be balanced against the richness of ends**.

God can do everything that is possible, but he will do only what is best. So God has created the best possible, the most perfect world, in that the greatest possible diversity of phenomena are governed by the smallest possible set of ideas. God simultaneously maximizes the richness and diversity of the world and minimizes the complexity of the decrees/hypotheses/mathematical laws, that determine this world.

All simple substances or **Monads** are **incorporeal automata**.

Corporeal mass, which is thought to have something over and above simple substances, is not a substance but a phenomenon resulting from simple substances (Monads) which alone have unity and absolute reality. Things which are uniform and contain no variety are never anything but abstractions. Extension and motion, are not substances, but true phenomena. "Time and space are not things, but orders of things."

There is perfect **was pre-established harmony** from the outset between the system of final causes and that of efficient causes, e.g., between the perceptions of the monad and the movements of bodies.

The Monads with *perceptions* that are more distinct and accompanied by *memory* are to be called 'souls'. Our knowledge of necessary truths, and our grasp of the abstractions they involve, raise us to the level of acts of *awareness/consciousness/reflection*, which make each of us aware of the thing that is called *I*, and let us have thoughts about this or that thing in us.

Each monad is a living substance. Every living substance is made up of smaller living substances which in their turn are made up of still smaller ones, and so on down to infinity. There are infinite levels of life among monads, some of which are more or less dominant over others. God is the greatest Monad.

There is interconnection of monads between each other. Each monad has relational properties that express all the others, so that each monad is a perpetual living mirror of the universe.

"The Universe is only the collection of a certain kind of compossibles; and the actual Universe is the collection of all existent possibles, i.e. of those which form the richest compound. And as there are different combinations of possibles, some better than others, there are many possible Universes, each collection of compossibles making one of them."

§1.2 Epistemology

Any finite set of points on a piece of graph paper always seems to follow a law, because there are infinite lines passing through those very points. But there is a law only if the line is simple.

The same effect can have several causes. Hence no firm demonstration can be made from the success of hypotheses. Yet the number of phenomena which are happily explained by a given hypothesis may be so great that it must be taken as morally certain. These hypotheses are to be presented, in the interim, in place of the true causes. It is also useful to apply less perfect hypotheses as substitutes for truth until a better one occurs, that is, one which explains the same phenomena more happily or more phenomena with equal felicity. There is no danger in this if we carefully distinguish the certain from the probable.

Leibniz invented binary arithmetic, and since the world is mechanical, he conceived the world mathematically as a peculiar "progression" of 0s and 1s. He believed that "One suffices to derive all out of nothing."

Leibniz believed that we can invent a universal calculus, to collect and reduce all human knowledge to numbers, then formalize it so everything one could ever want to know could be derived by essentially mathematical means. By forming all possible combinations of statements, one could systematically generate all possible knowledge. When it comes to finding an ultimate model expressed in the right language for the universe, we get to find a precise, exact, representation of the universe, with no approximations. So that, in a sense, we successfully reduce all of physics to mathematics. So that we would have, in a sense, achieved the great goal—of turning every question about the world into a question about calculus.

> *"Our characteristic will reduce the whole to numbers, so that reasons can also be weighed, as if by a kind of statics. For probabilities, too, will be treated in this calculation and demonstration, since one can always estimate which of the given circumstances will more probably occur."*
>
> — *Leibniz*

§2 Digital Philosophy—Lovely and Trustworthy :-)

§2.1 Feng Ye's Naturalism

Methodological naturalism is the inductive method of science. There is no higher tribunal for truth than natural science itself. There is no better method than the scientific method for judging the claims of science, and there is neither any need nor any place for a "first philosophy", such as (abstract) metaphysics or epistemology, that could stand behind and justify science or the scientific method. Science is the best way to explore the processes of the universe and that those processes are what modern science is striving to understand.

In Popper's view, the advance of scientific knowledge is an evolutionary process characterised by

$$PS_1 \longrightarrow TT_1 \longrightarrow EE_1 \longrightarrow PS_2 \longrightarrow \cdots$$

In response to a given problem situation (PS_1), a number of competing conjectures, or tentative theories (TT), are systematically subjected to the most rigorous attempts at falsification possible. This process, error elimination (EE), performs a similar function for science that natural selection performs for biological evolution. Theories that better survive the process of refutation are not more true, but rather, more "fit"—in other words, more applicable to the problem situation at hand (PS_1). Consequently, just as a species' biological fitness does not ensure continued survival, neither does rigorous testing protect a scientific theory from refutation in the future. For Popper, it is in the interplay between the tentative theories (conjectures) and error elimination (refutation) that scientific knowledge advances toward greater and greater problems; in a process very much akin to the interplay between genetic variation and natural selection. Although single observational events may prove hypotheses wrong, no finite sequence of events can verify them correct. Thus induction is theoretically unjustifiable and becomes in practice the choice of the simplest generalization that resists falsification, on the ground that the simpler a hypothesis is the easier it is to falsify it. But falsifiability is as subjective as simplicity. And sometimes a complex theory with fixed parameters is as easy to falsify as a simple theory. Maybe we should add that the tentative theories TT_i should be well-ordered according to some aesthetic criterion, e.g., the simplicity criterion (Leibniz would like) or generality/universality criterion, or to some utilitarian criterion if under some utility request. In the absence of compelling experimental or observational results, deciding which tentative theory/hypothesis/mathematics should be taken seriously is as much art as it is science.

Ontological naturalism is a philosophical worldview and belief system that holds that there is nothing but natural elements, principles, and relations of the kind studied by the natural sciences, i.e., those required to understand our physical environment by mathematical modeling. All are physical objects—there is nothing over and above the physical world—including we humans ourselves, and there is no non-physical 'subject', and therefore there are no abstract objects seen from the point of view of a 'subject' facing an 'external world'. It is 'objects with no subject'.

In the first chapter of [27], Feng Ye claimed that 'naturalism' is a modest, cautious and down-to-earth attitude. We are not sure if there are immaterial minds that can somehow 'grasp' mathematical concepts by some sort of intuition, and we do not know how to start studying such minds if we hypothesize them. However, we are quite sure that there are brains and that the neural net-works in human brains can do pattern recognition, language parsing, memory association, concept formation, logical inference, and so on and so forth. Therefore, why don't we start from our mainstream scientific description of human cognitive activities and investigate, from the logical and philosophical point of view, whether or not this is sufficient to give an account of human mathematical practices?

Then Feng Ye tried to naturalized the applicability of mathematics—to answer the following question:

Puzzle 1 (**Applicability Puzzle**)**.** ***Why is the applicability of classical mathematics possible?***

Feng Ye[27] convincingly justified 'The Conjecture of Finitism': Strict finitism is in principle sufficient for formulating current scientific theories about natural phenomena above the Planck scale and for conducting proofs and calculations in those theories. In other word, mathematical theorems about infinite mathematical entities are not really among the logically minimum premises implying our scientific conclusions about finite physical things, and our scientific conclusions are literally true of finite physical things because they logically follow from literally true premises about finite physical things. Applying mathematics in strict finitism is essentially using a computational device (including a brain) to simulate other physical entities and their properties.

The conventional approach holds that a theory of mathematical physics can be broken down into (i) a mathematical structure $[m]$, (ii) an empirical domain $[r]$ and (iii) a set of correspondence rules $[b]$ which bridge parts of the mathematical structure with parts of the empirical domain.

Let Γ_r be the collection of realistic premises in a specific application instance; let Γ_m be the collection of mathematical premises, including the premises expressing scientific laws and the mathematical axioms of classical mathematics; and let Γ_b be the collection of bridging postulations in that application. The application is then a purely logical inference

$$\Gamma_r \cup \Gamma_m \cup \Gamma_b \vdash \phi$$

from these premises to a realistic sentence ϕ as the realistic conclusion.

The Logical Problem of Applicability: In a scientifically valid application, assuming that $\mathfrak{A}_r \vDash \Gamma_r$, why does $\Gamma_r \cup \Gamma_m \cup \Gamma_b \vdash \phi$ imply $\mathfrak{A}_r \vDash \phi$, for ϕ is scientifically meaningful?

Feng Ye solved this problem by demonstrating that we only need $\Gamma_r' \subset \Gamma_m \cup \Gamma_b$ to get $\Gamma_r \cup \Gamma_r' \vdash \phi$ rather than $\Gamma_r \cup \Gamma_m \cup \Gamma_b \vdash \phi$. So the logical explanation of applicability goes like this:

$$\mathfrak{A}_r \vDash \phi \text{ because } \mathfrak{A}_r \vDash \Gamma_r \cup \Gamma_r' \text{ and } \Gamma_r \cup \Gamma_r' \vdash \phi$$

§2.2 Digital Philosophy

If a formal system describes a mathematical structure whose existence is guaranteed by some high-level theory, the latter is said to be a model of the former. But what is the real physical model $\mathfrak{A}_r$?

Actually, models are not god-given perfect structures, but are our constructions within our theory itself. No substantial part of the universe is so simple that it can be grasped and controlled without abstraction. We know our world is very complex and hard to understand, in order to make things easier to grasp, reality is generally broken down into bitesize chunks. We abstract our models from these chunks under consideration with similar but simpler structures with the help of our theories.

Normally we have to be content with making models that are approximations. Then we have to argue about whether we've managed to capture all the features that are essential for some particular purpose or not. Let the model approach asymptotically the complexity of the original chunk. It will tend to become identical with that chunk. As a limit it will become that chunk itself. The deepest description of the universe should not require concepts whose meaning relies on human experience or interpretation. Reality transcends our existence and so shouldn't, in any fundamental way, depend on ideas of our making. Let's imagine what the ultimate model looks like. The ultimate model of the world must be the world itself. In other word, should the ultimate model thoroughly realized its purpose, the world could be grasped in its entirety and the model would be unnecessary (or the world would be unnecessary).

But we do not have a right semantic language at hand. There is usually a tradeoff between tractability and comprehension on one hand, and accuracy on another. Any useful semantic description of a phenomenon will lie on a Pareto front trading tractability with accuracy.

Leibniz wanted to invent a universal calculus, to collect and reduce all human knowledge to numbers, then formalize it so everything one could ever want to know could be derived by essentially mathematical means. When it comes to finding an ultimate model expressed in the right language for the universe, we get to find a precise, exact, representation of the universe, with no approximations. So that, in a sense, we could reduce all of physics to mathematics.

Wittgenstein was concerned with the conditions for a logically perfect language. In order that a certain sentence should assert a certain fact there must, however the language may be constructed, be something isomorphic between the structure of the sentence and the structure of the fact. That which has to be in common between the sentence and the fact cannot be itself in turn said in language. It can only be shown, not said, for whatever we may say will still need to have the same structure. That is to say, there is a perfect 1-1 correspondence between the world and the language. No redundancy should appear in the language.

What if it is possible to give a thorough description of the physical reality involving no presupposed parameters? If so, our description of the reality would have to be completely abstract, forcing any words or other symbols used to denote them to be mere labels with no preconceived meanings whatsoever, which leads us to Tegmark's 'Mathematical Universe Hypothesis'(MUH): Our physical reality is a mathematical structure. Mathematics itself already is real; it doesn't require instantiation. Some collections of mathematical equations are different universes. But note that not all of the recursively axiomatizable formal systems characterize universes. In mathematics we end up with recursively axiomatizable formal systems, and they are trying to sculpt things, to ensure that they're really talking about the thing that we originally imagined we were talking about, like integers. But what we know from Compactness theorem and Lowenheim-Skolem theorem and so on, is that that kind of sculpting can never really work. We can never really use this method to uniquely characterize a structure up to isomorphism.

MUH explains the utility of mathematics for describing the physical world as a natural consequence of the fact that the latter is a mathematical structure, and we are simply uncovering this bit by bit. The various approximations that constitute our current physics theories are successful because simple mathematical structures can provide good approximations of certain aspects of more complex mathematical structures. In other words, our successful theories are not mathematics approximating physics, but mathematics approximating mathematics. What we have to explain is not the applicability of mathematics, but the emergence of physical phenomena.

If the MUH is correct, then (ii) and (iii) are redundant in the sense that they can, at least in principle, be derived from (i).

Assume $\mathfrak{A}_r$ is the ultimate model of our world, then the task of natural science is to obtain a thorough understanding and control of $\mathfrak{A}_r$, i.e. to grasp Theory of Everything $Th(\mathfrak{A}_r)$. But, is it possible? Is our world knowable? First of all, what is the meaning of 'knowable'?

For mathematical structures, formal systems, and computations alike, there is a subset with an attractive property (being defined, decidable and computable, respectively).

We assume that anything that can be called knowable/understandable/comprehensible can be simulated by some mechanical device to arbitrary precision.

Knowability = Decidability/Computability

Maybe Hilbert would like the assumption.

> *"There is no ignorabimus in mathematics."*
>
> *"We must know; We will know."*

— *Hilbert*

Gödel proved that $Th(\mathfrak{N})$ is unknowable: $Th(\mathfrak{N})$ is undecidable, i.e. $Th(\mathfrak{N})$ is complete but not r.e.; $Th(PA)$ is r.e. but not complete.

For a finite structure $\mathfrak{A}$ in a finite language, $Th(\mathfrak{A})$ is decidable.

If our universe is finite and discrete, then we can say it is comprehensible.

But why is our universe finite and discrete? Or we can ask ourselves 'why our universe is knowable?', or:

Puzzle 2 (**Knowability Puzzle**). ***Why is natural science possible?***

According to Leibniz, God has created the best possible world balancing the simplicity of means against the richness of ends. That is why natural science is possible!

> *"God does not play dice."*
>
> *"God always takes the simplest way."*
>
> *"Subtle is the Lord, however he is not malicious."*
>
> *"The most incomprehensible thing about the world is that it is comprehensible."*
>
> — *Einstein*

Since for any $\mathfrak{A}$, $Th(\mathfrak{A})$ is complete, to make sure it is knowable, it only needs to be r.e.

Hypothesis 1 (Knowability Hypothesis)**.**
For any universe $\mathfrak{A}$, $Th(\mathfrak{A})$ is r.e.

According to the Knowability Hypothesis, the mathematical structure that is our universe is computable and hence well-defined in the sense that all its properties and relations can be computed/decided. There are thus no physical aspects of our universe that are incomputable or undecidable.

Hence, $\mathcal{M}_T := \{\mathfrak{A} : Th(\mathfrak{A}) \text{ is r.e.}\}$ is the class of universes knowable to us.

For a complete axiomatizable theory (in a reasonable language) is decidable, $Th(\mathfrak{A})$ can be encoded to a computable real number $0.x$. Then why not generalize $\mathcal{M}_T$ a little bit?

$$\mathcal{M}_D := \{p : \exists x\, (U(p) = x)\}$$

where U is the universal monotone Turing machine.

$\mathcal{M}_D$ corresponds to Schmidhuber's multiverse.

Loosely(very loosely),

$$\mathcal{M}_T \subset \mathcal{M}_D \in \mathcal{M}_D \setminus \mathcal{M}_T$$

It's easy to design a dovetailer program to generate all possible computable universes.

Within the universal dovetailer's output is a description of a computer running another dovetailer, and so on ad infinitum. It is entirely possible that we do exist as a description within the output of a dovetailer, which itself may well be described in the output of another dovetailer.

The entirety of a simulation, run from start to finish, is itself a collection of mathematical relations. Thus, if one believes that all mathematics is real, so is this collection. In turn, from this perspective there's no need to actually run any computer simulations since the mathematical relations each world produce are already real. The computability of a universe should be evaluated by examining the computability of the mathematical relations that define its entire history, whether or not these relations describe the unfolding of the simulation through time.

The theory of universal Turing machines is a mathematical theory, and a universal dovetailer is a formal specification. But we can't find $\mathcal{M}_D$ within $\mathcal{M}_T$ because of Gödel incompleteness theorem: for any recursively axiomatized, consistent formal system that is strong enough to represent primitive recursive functions, it is incomplete!

Now we can take our universe as a giant computer program, just like the virtual reality in *Matrix*. Our minds can be perfectly simulated on some type of Turing machine, the appearance of a physical material reality experienced by the mind follows, without the need of any concrete computer.

The diagonal method shows that the logically consistent 'universes' are conceivable, but cannot be realized in virtual reality.

Maybe physical models, mathematical structures, formal systems, and computations are simply different aspects of one underlying transcendent structure (Multiverse) whose nature we still do not fully understand.

Moreover, many people think that we were shaped by evolution to find patterns in the environment; the better we could do that, the better we could predict how to find the next meal. Mathematics, the language of pattern, emerged from our biological fitness. And with that language, we've been able to systematize the search for new patterns, going well beyond those relevant for mere survival. But mathematics, like any of the tools we developed and utilized through the ages, is a human invention.

Then, why should we reduce the (concrete) physical reality to (abstract) computation, to programs, to a sequence of 0s and 1s? Why is the incorporeal computation is more fundamental than our tangible sense data? Why do we think 'hardware' is made of 'software'? Because in this way, we can solve Applicability Puzzle1 and Knowability Puzzle2 at a draught, and it is a good hypothesis, for we have strong faith in Church-Turing Thesis.

Thesis 1 (Church-Turing Thesis).

effective calculable = recursive = finite definable = Herbrand-Gödel computable = representable in any consistent formal system extending $\mathcal{R}$[1] *= λ-definable = Turing Computable = flowchart (or 'while') computable = neural network with a tape computable = Conway's 'game of life' = Post/Markov/McCarthy/Kolmogorov and Uspensky ...*

The notion of recursiveness is very stable. The class of recursive functions is not sensitive to changes in the formal systems considered to represent its functions: that is, the same functions are representable in any consistent formal system having a least minimal power, independently of the system strength. Even in vast classes of recursive transfinite progressions of formal systems, only recursive functions are representable. Thus the notion is absolute in a certainly astonishing way.

It is plausible that any possible behavior of a discrete physical system (including the brain) (according to present day physical theory) is recursive.

David Deutsch: "Every finitely realizable physical system can be perfectly simulated by a universal model computing machine operating by finite means."

On one hand, we can take our universe as a giant program, and natural science as a theory of computation. On the other hand, our computers are physical objects, and computations are physical processes. What computers can or cannot compute is determined by the laws of physics. However, many of the discrete physical systems, for example, our PCs and our brains, can be taken as universal Turing machines if supplemented with infinite memory. So there is some comprehensive self-similarity in our world: it is possible to build a virtual-reality generator whose repertoire includes every physically possible environment. It's much easier to program a computer to generate all possible computable universes than it is to program individual computers to generate them one by one. So a single, buildable object

[1] $\mathscr{L}_{\mathcal{R}} = \{0, S, +, \cdot, <\}$

A1. $\neg(\underline{x} = \underline{y})$ for $x \neq y$

A2. $x < \underline{n} \vee x = \underline{n} \vee \underline{n} < x$

A3. $\neg(x < \underline{0})$

A4. $x < \underline{n+1} \leftrightarrow x = \underline{0} \vee \ldots \vee x = \underline{n}$

A5. $\underline{x} + \underline{y} = \underline{x+y}$

A6. $\underline{x} \cdot \underline{y} = \underline{x \cdot y}$

$\mathcal{R}$ is not finitely axiomatizable, and any closed theorem of $\mathcal{R}$ is true in some finite model. Robinson arithmetic $\mathcal{Q}$ is an extention of $\mathcal{R}$, and thus provides an example of a (minimal) finitely axiomatized theory in which every computable function is representable.

can mimic all the behaviors and responses of any other possible object or process. This is what makes reality comprehensible. And this is also consistent with Plato's idea—'learning = recollection'. Computations can generate the class of all theorems of recursively axiomatizable formal systems. We can use mathematics to help us to understand our own universe.

§2.3 Leibniz's Universe and How to Live in Leibniz's Universe

> *"The Universe is only the collection of a certain kind of compossibles; and the actual Universe is the collection of all existent possibles, i.e. of those which form the richest compound. And as there are different combinations of possibles, some better than others, there are many possible Universes, each collection of compossibles making one of them."*
>
> — *Leibniz*

Every program from $\mathcal{M}_D$ determines an universe. But which one is the best possible world? The one with the richest phenomena governed by the simplest hypotheses. To quantify the simplicity, we can resort to Kolmogorov Complexity. But we do not know how to quantify the richness. Since a monotone Turing machine may fail to halt, many of $\{U(p) : p \in \mathcal{M}_D\}$ are infinite in length.

Leibniz's universe consists of Monads. Each monad is a program. So the universe is a pool of programs interacting with each other. One can compute the output of the others, simulate the others, or itself.

Some combination of programs makes itself a well-defined program. Leibniz's universe must be a giant program consisting of lots of programs which can be integrated together, and some of them can generate all of the programs, i.e. the $\mathcal{M}_D$ itself. An 'agent' living in such a universe can distinguished itself with the others. How can such an 'agent' discover the truth of the universe?

Theory of Everything

How to obtain deep understanding of our universe? The agent has to grasp the Theory of Everything (ToE) that includes the knowledge of herself.

One may define the best Theory of Everything (ToE) of an observer with experience h as

$$\langle p^*, q^* \rangle [h] := \arg\min_{p,q} \{ |\langle p, q \rangle| : U(p, U(\langle p, q \rangle, \epsilon)) = h* \}$$

Leibniz's universe is deterministic. What if the universe can be stochastic?

$$\mathcal{M}_U := \{ \rho : \mathcal{X}^* \cup \mathcal{X}^\infty \to [0,1] : \rho \text{ is an enumerable semimeasure} \}$$

A theory that predicts universe ω with probability $\mu(\omega)$ and experience h in universe ω with probability $\pi(h|\omega)$, induces a probability distribution $P(h) := \sum_{\omega} \pi(h|\omega)\mu(\omega)$. The observed noise can then be coded in $-\log_2 P(h)$ bits.

A best computable probabilistic ToE with experience h is the one (π, μ) which minimizes

$$|p_\pi| + |p_\mu| - \log_2 P(h)$$

where p_π and p_μ are the programs that computes π and μ respectively.

Anthropic Principle requires $\pi(h|\omega) = 1$, which leads to MDL if we ignore $|p_\pi|$.

$$|p_\mu| - \log_2 \mu(h)$$

Note that we are the inseparable part of the universe, and our experience h is the result of the interaction between the 'ego' and 'environment', so π must be implicitly involved in μ in some way, just like the deterministic ToE $\langle p, q\rangle$. Hence the best computable probabilistic ToE with experience h is

$$\arg\min_{\frac{\pi}{\mu}}\{|p_{\frac{\pi}{\mu}}| - \log_2 \tfrac{\pi}{\mu}(h)\} = \arg\max_{\frac{\pi}{\mu}}\{2^{-|p_{\frac{\pi}{\mu}}|} \cdot \tfrac{\pi}{\mu}(h)\} = \arg\max_{\frac{\pi}{\mu}}\{P\left[\tfrac{\pi}{\mu}|h\right]\}$$

So far, we have defined a good guess of ToE according to the inductive method of science, is there any better way to predict the future?

Probability in Leibniz's Universe

> *"Our characteristic will reduce the whole to numbers, so that reasons can also be weighed, as if by a kind of statics. For probabilities, too, will be treated in this calculation and demonstration, since one can always estimate which of the given circumstances will more probably occur."*
>
> — *Leibniz*

Our beliefs and hence probabilities are a result of our personal history. To be able to update beliefs consistently we must first decide on the set of all explanations that may be possible. In order to find the true governing process behind our entire reality we consider Leibniz's all possible universes ($\mathcal{M}_D$) in a certain sense. The actual universe is just one of a large number of possible universes. Each universe is in one of possible states; the probability assigned to each state is then the proportion of the possible universes in which that state is attained. Each new measurement eliminates some fraction of the universes in a given state, depending on how likely or unlikely that state was to actually produce that measurement; the surviving universes then have a new posterior probability distribution, which is related to the prior distribution by Bayes' formula.

Definition 1 (**Universal Probability**).

$$M(x) := \sum_{p:U(p)=x*} 2^{-|p|}$$

where U is a universal monotone Turing machine.

It can be regarded as the limit of the relative frequency of the consistent possible worlds over all possible worlds (complying with the frequentist interpretation):

$$\begin{aligned} M(x) &= \sum_{p} 2^{-|p|} [\![U(p) = x*]\!] \\ &= \lim_{n\to\infty} \frac{\sum_{p:|p|\le n} 2^{n-|p|} [\![U(p) = x*]\!]}{2^n} \\ &\approx \lim_{n\to\infty} \frac{|\{p \in \chi^n : U(p) = x*\}|}{2^n} \end{aligned}$$

It means that $M(x)$ is the frequentist probability that the output of a universal monotone Turing machine U starts with x when provided with uniform random noise (fair coin flips) on the input tape.

The benevolence of God is represented in the way he plays dice. God does not play dice directly with us, but plays dice indirectly through some Universal Turing machine to offer us the freedom to realize any possible regular world. And God's dice is absolutely fair, which means God never play tricks. God offers us the fullest freedom to chose the most perfect world, that is to say, the one which is at the same time the simplest in hypothesis and the richest in phenomena.

In $\mathcal{M}_U$ we can define the universal bayes mixture.

Definition 2 (**Universal Bayes Mixture**).

$$\xi_U(x) := \sum_{\nu\in\mathcal{M}_U} w_\nu^U \nu(x)$$

where $w_\nu^U := 2^{-K(\nu)}$ is the universal prior.

It can be shown that $\xi_U \in \mathcal{M}_U$.

If we let $\mathcal{M}_U$ be the class of all computable distributions, then ξ_U would be not lower semi-computable, since the class of all computable distributions is not recursively enumerable, that is why we "slightly" extend the class to include also enumerable semimeasures.

Obviously, $M(x)$ can also be regarded as a $w_{\nu_p} := 2^{-|p|}$-weighted mixture over all computable deterministic environments ν_p ($\nu_p(x) = 1$ if $U(p) = x*$ and 0 otherwise). We call $2^{-|p|}$ the Solomonoff prior.

Lemma 1. *For every $\nu \in \mathcal{M}_U$ there exists some monotone Turing machine T such that*

$$\nu(x) = \sum_{p:T(p)=x*} 2^{-|p|} \quad \textit{and} \quad K(\nu) \stackrel{+}{=} |\langle T \rangle|$$

where $T(p) = U(\langle T \rangle p)$.

Theorem 1 (**Deterministic Representation of Bayesian Mixture).**

$$M(x) \stackrel{\times}{=} \xi_U(x)$$

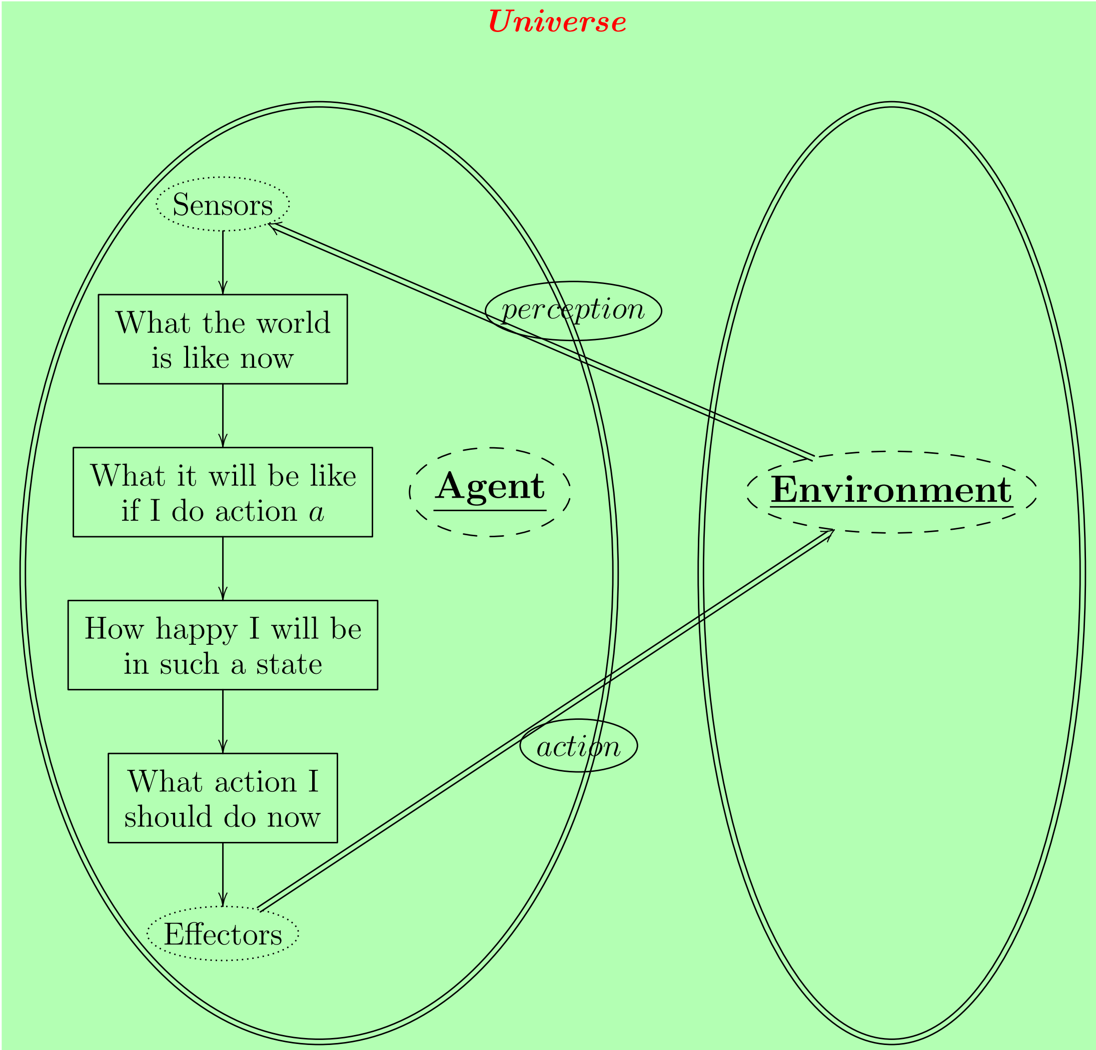
In the following we'll show that: Intelligence of an 'agent' is an Perfect Bayesian Equilibrium of the Incomplete Information game with observable actions played against imaginary possible worlds ($\mathcal{M}_U$) only if our subjective belief system is the universal prior; or equivalently, Intelligence can be seen as an Ex Post Equilibrium of the Incomplete Information game with observable actions played against imaginary possible worlds ($\mathcal{M}_U$) only if we pretend that the true environment is the universal mixture environment; or equivalently, Intelligence can also be regarded as an Perfect Bayesian Equilibrium of the Incomplete Information game with observable actions played against Leibniz's possible worlds ($\mathcal{M}_D$) only if our subjective belief system is the Solomonoff prior.

§3 From Leibniz to AIXI via Two Types of Games

§3.1 Agents in Known (Probabilistic) Environment

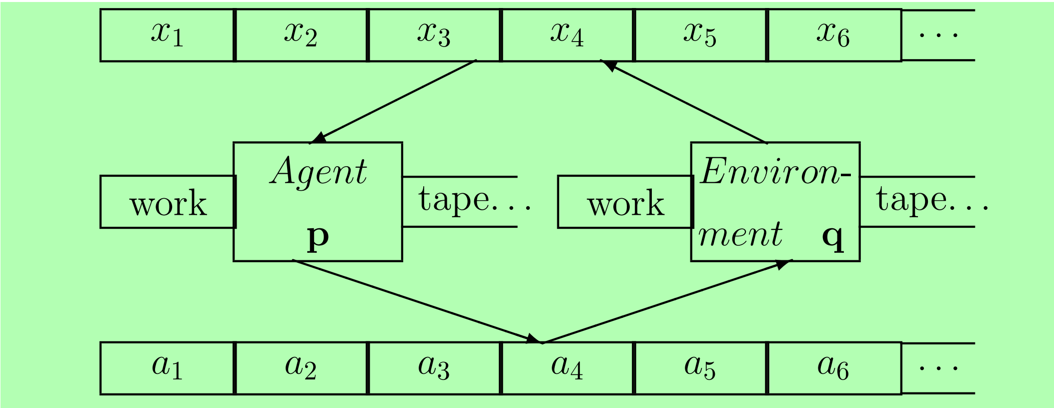
"The environment is everything that isn't me."

— *Einstein*



Consider an agent that exists within some unknown environment. The agent interacts with the environment in cycles. In each cycle, the agent executes an action and in turn receives a perception. The only information available to the agent is its history of previous interactions. The environment reacts to the agent's action and leads to a new perception (input) to the agent determined by a deterministic procedure or some probability distribution which depends on the history. The problem is to construct an agent that, over time, collects as much utility as possible from the (unknown) environment.

If input and output are represented by strings, a deterministic procedure can be modeled by a Turing machine p. p is called the policy of the agent, which determines the (re)action to a perception. If the environment is also computable it might be modeled by a Turing machine q as well. The interaction of the agent with the environment can be illustrated as follows:



p as well as q have unidirectional input and output tapes and bidirectional work tapes. What entangles the agent with the environment, is the fact that the upper tape serves as input tape for p, as well as output tape for q, and that the lower tape serves as output tape for p as well as input tape for q. Further, the reading head must always be left of the writing head, i.e. the symbols must first be written, before they are read. The heads move in the following way. In the k^{th} cycle p writes a_k, q reads a_k, q writes x_k, p reads x_k, followed by the $(k+1)^{th}$ cycle and so on. This continues ad infinitum or for a finite number of cycles. The whole process starts with the first cycle, all heads on tape start and work tapes being empty. We want to call Turing machines behaving in this way *chronological Turing machines.*

Let us define for the chronological Turing machine p a partial function also named $p: \mathcal{X}^* \to \mathcal{A}^*$ with $a_{1:k} = p(x_{<k})$, where $a_{1:k}$ is the output of Turing Machine p on input $x_{<k}$ in cycle k, i.e. where p has read up to x_{k-1} but no further. In an analogous way, we define $q: \mathcal{A}^* \to \mathcal{X}^*$ with $x_{1:k} = q(a_{1:k})$. Conversely, for every partial recursive chronological function we can define a corresponding chronological Turing machine.

§3.1.1 From Bayesian Game Γ_1 to AIμ

Life is the sum of all our choices. To live is to choose—to exercise the gift of free will; but to choose well, we need to foresee the future; to foresee the future, we need to estimate all of the possible futures; and to estimate the possible futures, we need to analyze the history.

An agent's policy p should be entirely described by:

- the finite set of possible actions $\mathcal{A}$ and perceptions $\mathcal{X}$;
- its utility function $u: \mathbb{N} \times \mathcal{H} \to \mathbb{R}$, where $\mathcal{H} := (\mathcal{A} \times \mathcal{X})^* \cup (\mathcal{A} \times \mathcal{X})^* \times \mathcal{A}$, and $|h|$ denotes the length of the history $h \in \mathcal{H}$, and we denote $\mathcal{H}^t := \{h \in \mathcal{H} : |h| = t\}$ and $\mathcal{H}^\infty := (\mathcal{A} \times \mathcal{X})^\infty$.
- its temporal discount function $d: \mathbb{N}^2 \to \mathbb{R}$ s.t. $\forall t \forall k \geq t : \Gamma_{tk} := \sum_{i=k}^{\infty} d(t,i) < \infty$;

- its prior knowledge $\mu\colon \mathcal{Q} \to [0,1]$ of the environments $\mathcal{Q}$.

If $h = ax_{1:t}$, we write $x(h) = q(a(h))$, where $x(h) = (x_1, \ldots, x_t)$ and $a(h) = (a_1, \ldots, a_t)$, to denote the output of $q \in \mathcal{Q}$ producing x_i in response to the actions a_i for $1 \leq i \leq t$, which means $q \rightsquigarrow h$.

$$\mathcal{Q}_h := \{q \in \mathcal{Q} : x(h) = q(a(h))\}$$

Similarly, if $h = ax_{<t}a_t$, we write $a(h) = p(x(h))$ to denote the output of $p \in \mathcal{P}$ producing $a(h)$ in response to the perceptions $x(h)$ for $1 \leq i \leq t$, which means $p \rightsquigarrow h$.

$$\mathcal{P}_h := \{p \in \mathcal{P} : a(h) = p(x(h))\}$$

Each (agent,environment) pair (p, q) produces a unique I/O sequence $\omega^{pq} = a_1^{pq} x_1^{pq} a_2^{pq} x_2^{pq} \ldots$ in chronological order.

Assume the environment model is the other rational player with constant utility

$$V_t^{qp}(h) := \begin{cases} 1 & \text{if } q \in \mathcal{Q}_h \\ 0 & \text{otherwise} \end{cases}$$

Let

$$V_t^{pq}(ax_{<k}) := \sum_{i=k}^{\infty} d(t,i) u(t, ax_{1:i}^{pq})$$

be the future total utility that the agent p receives from the environment q from the cycle k at time t.

We see that the agent-environment configuration satisfies all criteria of a strategic form game $\langle \{a, e\}, (\mathcal{P}, \mathcal{Q}), v \rangle$(or an extensive form game with perfect information $\langle \{a, e\}, P, \mathcal{H}, v \rangle$).

Nash Equilibrium

Definition 3 (Nash Equilibrium)**.** *A (pure strategy) Nash equilibrium of a strategic form game is a strategy profile* (p^*, q^*) *such that*

$$p^* \in R_a^B(q^*)$$

$$q^* \in R_e^B(p^*)$$

where

$$R_a^B(q) := \left\{ p \in \mathcal{P}_h : \forall p' \in \mathcal{P}_h \left[V_t^{pq}(h) \geq V_t^{p'q}(h) \right] \right\}$$
$$R_e^B(p) := \left\{ q \in \mathcal{Q}_h : \forall q' \in \mathcal{Q}_h \left[V_t^{qp}(h) \geq V_t^{q'p}(h) \right] \right\}$$

Hence, a Nash equilibrium is a strategy profile (p^*, q^*) such that no player a/e can profit by unilaterally deviating from his strategy p^*/q^*, assuming every other player e/a follows his strategy q^*/p^*.

Given a specific deterministic environment q, the best strategy of the agent at time t with possible future history h is the one which maximizes $V_t^{pq}(h)$,

$$p^* \in \arg\max_{p \in \mathcal{P}_h} V_t^{pq}(h)$$

Since the environment's utility is invariant, the Nash equilibriums are

$$\{(p,q) : p \in \arg\max_{p \in \mathcal{P}_h} V_t^{pq}(h)\}$$

However, there are too many Nash equilibriums and the environment is unknown. And there exists no dominant equilibrium (p^*, q^*) such that

$$\forall p \in \mathcal{P}_h \forall q \in \mathcal{Q}_h \left[V_t^{p^* q}(h) \geq V_t^{pq}(h)\right]$$

$$\forall p \in \mathcal{P}_h \forall q \in \mathcal{Q}_h \left[V_t^{q^* p}(h) \geq V_t^{qp}(h)\right]$$

Harsanyi Transformation The problem is: *how to live with uncertainty? how to live on almost nothing but beliefs?*

We should turn to *Harsanyi transformation* for help. Harsanyi introduces to the game the notion of *nature's choice* or *God's choice*, according to which we can convert a game with incomplete information into a game with complete yet imperfect information. He proposed treating a player who has different payoffs under different circumstances as a player of different types. The game is then modeled as though 'nature' moves first and chooses that player's type. In this kind of game a player must form beliefs about the strategy that an opponent will play and the player must also form some belief about the type of game she is playing.

Strategy for Known Prior Probability Let us now weaken our assumptions by replacing the environment q with a probability distribution $\mu(q) : \mathcal{Q} \to [0,1]$.

Assume here that the probability $\mu(q)$ is the agent's subjective prior belief that q is a true model of the environment.

The two functions, u and w, allow the agent at time t to put a value $V_t^{pq}(h) = \sum_{i=k}^{\infty} d(t,i) u(t, h_{1:i}^{pq})$ on each possible history h based on what futures are possible given a particular action set.

From Bayesian Game Γ_1 to AIμ

*"It will have perfect **power**, **knowledge**, and **will**; that is to say, it will have **omnipotence**, **omniscience**, and **sovereign goodness**."*

— *Leibniz*

For the *Bayesian extensive game with observable actions* $\Gamma_1 = \langle N, P, \mathcal{H}, (\Theta_i)_{i\in N}, (p_i)_{i\in N}, (u_i)_{i\in N}\rangle$, where $N = \{a, e\}$, $\Theta_a = \mathcal{P}$, $\Theta_e = \mathcal{Q}$, $p_e(q) = \mu(q)$, the behavioural strategy $\rho_i(\theta_i)(h)(a)$ of Γ_1 becomes

$$\rho_a(p)(h)(a) = \begin{cases} 1 & \text{if } p \in \mathcal{P}_{ha} \\ 0 & \text{otherwise} \end{cases}$$

$$\rho_e(q)(h)(a) = \begin{cases} 1 & \text{if } q \in \mathcal{Q}_{ha} \\ 0 & \text{otherwise} \end{cases}$$

and the belief system $\mu_i(h)(\theta_i)$ can be proved to be

$$\mu_e(h)(q) = \begin{cases} \frac{\mu(q)}{\sum\limits_{q'\in\mathcal{Q}_h} \mu(q')} & \text{if } q \in \mathcal{Q}_h \\ 0 & \text{otherwise} \end{cases}$$

At time t, the players' payoff function

$$u_a|_h(h', (p, q)) := \sum_{i=|h|}^{\infty} d(t, i) u(t, hh'_{1:i})$$

$$u_e|_h(h', (p, q)) := \begin{cases} 1 & \text{if } q \in \mathcal{Q}_{hh'} \\ 0 & \text{otherwise} \end{cases}$$

At time t, with type p, the player a's ex post expected utility with possible future history h is

$$V_t^{pq}(h) = \begin{cases} \sum\limits_{i=|h|}^{\infty} d(t, i) u(t, h_{1:i}^{pq}) & \text{if } p \in \mathcal{P}_h \ \& \ q \in \mathcal{Q}_h \\ 0 & \text{otherwise} \end{cases}$$

where h^{pq} is the history uniquely determined by (p, q).

At time t, with type p, the player a's ex interim expected utility with possible future history h is

$$\begin{aligned} V_t^{p\mu}(h) &= \sum_{q\in\mathcal{Q}} V_t^{pq}(h) \mu_e(h)(q) \\ &= \sum_{q\in\mathcal{Q}_h} \left[\frac{\mu(q)}{\sum\limits_{q'\in\mathcal{Q}_h} \mu(q')} \right] V_t^{pq}(h) + \sum_{q\in\mathcal{Q}\setminus\mathcal{Q}_h} V_t^{pq}(h) \mu_e(h)(q) \end{aligned}$$

$$= \sum_{q \in \mathcal{Q}_h} \left[\frac{\mu(q)}{\sum_{q' \in \mathcal{Q}_h} \mu(q')} \right] V_t^{pq}(h)$$

Equilibrium are predictions of behaviour, to achieve the perfect bayesian equilibrium (ρ, μ), let

$$\rho_a(p)(h)(a') = \begin{cases} 1 & \text{if } p \in \arg\max_{p \in \mathcal{P}_h} V_t^{p\mu}(h) \ \& \ p(x(h)) = a(h)a' \\ 0 & \text{otherwise} \end{cases}$$

To make it clear, we define agent AIμ in functional form as follows.

Definition 4 (Agent AIμ in *Functional Form*).
Agent in functional form is defined as follows:

$$V_t^{pq}(ax_{<k}) := \sum_{i=k}^{\infty} d(t,i) u(t, ax_{1:i}^{pq}) \tag{Ex Post}$$

$$V_t^{p\mu}(ax_{<k}) := \frac{\sum_{q \in \mathcal{Q}_{ax_{<k}}} \mu(q) \cdot V_t^{pq}(ax_{<k})}{\sum_{q \in \mathcal{Q}_{ax_{<k}}} \mu(q)} \tag{Ex Interim}$$

$$p_t^{\mu} := \arg\max_{p \in \mathcal{P}_{ax_{<t}}} V_t^{p\mu}(ax_{<t})$$

At time t, the maximal achievable expected utility with possible future history $ax_{<k}$ in environment μ is

$$V_t^{*\mu}(ax_{<k}) := V_t^{p_t^{\mu}\mu}(ax_{<k})$$

p_t^{μ} depends on t and is used only in step t to determine a_t by $p_t^{\mu}(x_{<t}|a_{<t}) = a_{<t}a_t$. We can define agent

$$p^{\mu}(x_{<t}) := p_t^{\mu}(x_{<t}|p_{t-1}^{\mu}(x_{<t-1}| \ldots p_1^{\mu}(\epsilon) \ldots)) \tag{Perfect Bayesian}$$

or in one line

$$a_k^{*\mu} := \arg\max_{a_k} \max_{p: p(x_{<k}) = a_{<k}a_k} \frac{\sum_{q: q(a_{<k}) = x_{<k}} \left[\mu(q) \cdot \sum_{i=k}^{\infty} d(t,i) u(t, ax_{1:i}^{pq}) \right]}{\sum_{q: q(a_{<k}) = x_{<k}} \mu(q)}$$

§3.1.2 From Bayesian Game Γ_2 to AIμ

μ might be interpreted in two ways. Either the environment itself behaves stochastically defined by μ or the true environment is deterministic, but we only have subjective (probabilistic) information, of which environment being the true environment.

Now we assume that μ is known and describes the true (objective) stochastic behaviour of the environment.

Definition 5 (**Environment Model**).
An environment model ρ is a sequence of conditional probability functions $\{\rho_0, \rho_1, \rho_2, \ldots\}$, where $\rho_n: \mathcal{A}^n \to$ Density $(\mathcal{X}^n)$, that satisfies

$$\forall a_{1:n} \forall x_{<n} : \rho_{n-1}(x_{<n}|a_{<n}) = \sum_{x_n \in \mathcal{X}} \rho_n(x_{1:n}|a_{1:n}) \tag{3.1}$$

In the base case, we have $\rho_0(\epsilon|\epsilon) = 1$.

Equation(3.1), called the chronological condition, captures the natural constraint that action a_n has no effect on earlier perceptions $x_{<n}$. For convenience, we drop the index n in ρ_n from here onwards.

Similarly, we will in general call functions satisfying equation(3.2) *chronological semimeasures*, and *chronological measures* if "=" holds.

$$\forall a_{1:n} \forall x_{<n} : \sum_{x_n \in \mathcal{X}} \rho(x_{1:n}|a_{1:n}) \leq \rho(x_{<n}|a_{<n}) \text{ and } \rho(\epsilon|\epsilon) \leq 1 \tag{3.2}$$

Given an environment ρ, we define the predictive probability for $\forall a_{1:m} \forall x_{1:m} : \rho(x_{<n}|a_{<n}) > 0$

$$\rho(x_{n:m}|ax_{<n}a_{n:m}) := \frac{\rho(x_{1:m}|a_{1:m})}{\rho(x_{<n}|a_{<n})} \tag{3.3}$$

It follows that:

$$\rho(x_{1:n}|a_{1:n}) = \rho(x_1|a_1)\rho(x_2|ax_1a_2)\cdots\rho(x_n|ax_{<n}a_n) \tag{3.4}$$

The environment ρ can equivalently be written as a function $\rho: (\mathcal{A} \times \mathcal{X})^* \to [0,1]$ such that $\forall h \in (\mathcal{A} \times \mathcal{X})^* \forall a \in \mathcal{A} \Big(\sum_{x \in \mathcal{X}} \rho(x|ha) = 1 \Big)$ where we write $\rho(x|ha)$ for the function value of (h, a, x).

Utility, Policy and Value Functions The agent's goal is to accumulate as much utility as it can during its lifetime. More precisely, the agent seeks a *policy* that will allow it to maximise its expected future utility. The instantaneous utility values are assumed to be bounded.

The policy $\pi: (\mathcal{A} \times \mathcal{X})^* \times \mathcal{A} \to [0,1]$ and the environment $\mu: (\mathcal{A} \times \mathcal{X})^* \to [0,1]$ define the conditional probability of certain symbols given an interaction history: π defines the conditional probability over the actions, and μ of the perceptions. However, taken together they do define a measure over the interaction sequences that we will denote $\frac{\pi}{\mu}$.

$$\tfrac{\pi}{\mu}(ax_{1:t}) := \pi(a_1)\mu(x_1|a_1)\pi(a_2|ax_1)\mu(x_2|ax_1a_2)\cdots\pi(a_t|ax_{<t})\mu(x_t|ax_{<t}a_t)$$

From Bayesian Game Γ_2 to AIμ

For the *Bayesian extensive game with observable actions* $\Gamma_2 = \langle N, P, \mathcal{H}, (\Theta_i)_{i\in N}, (p_i)_{i\in N}, (u_i)_{i\in N}\rangle$, where $N = \{a, e\}$, $\Theta_a = \Theta_e = \mathcal{M}$ = "the set of all chronological measures", $p_e(\mu) = w_\epsilon^\mu$. We write $\mu_e(h)(\nu) := w_h^\nu$ as shorthand.

The behavioural strategy $\rho_i(\theta_i)$ of Γ_2 is π and μ.

The belief system can be proved to be

$$\begin{aligned}\mu_e(hx')(\mu) &= w^\mu_{hx'} \\ &= \frac{\mu(x'|h)w_h^\mu}{\sum_{\nu\in\mathcal{M}}\nu(x'|h)w_h^\nu} \\ &\overset{(a)}{=} \frac{\mu(x(h)x'|a(h))w_\epsilon^\mu}{\sum_{\nu\in\mathcal{M}}\nu(x(h)x'|a(h))w_\epsilon^\nu}\end{aligned}$$

where $\overset{(a)}{=}$ follows from mathematical induction.

So, the belief system $\mu_i(h)(\theta_i)$ is

$$\mu_e(h)(\mu) = \frac{\mu(x(h)|a(h))w_\epsilon^\mu}{\sum_{\nu\in\mathcal{M}}\nu(x(h)|a(h))w_\epsilon^\nu}$$

At time t, the players' payoff function

$$u_a|_h(h', (\pi,\mu)) := \sum_{i=|h|}^{\infty} d(t,i)u(t, hh'_{1:i})$$

$$u_e|_h(h', (\pi,\mu)) := \begin{cases}1 & \text{if } \tfrac{\pi}{\mu}(hh') > 0 \\ 0 & \text{otherwise}\end{cases}$$

At time t, with type π, the player a's ex post expected utility with possible future history h is

$$V_t^{\pi\mu}(h) := \mathbb{E}_{\frac{\pi}{\mu}}\left[u_a|_h(\cdot, (\pi,\mu)) \,\middle|\, h\right]$$

In other words:

Definition 6 (Ex Post Expected Future Utility).
After history of interaction $ax_{<t}$, the expected future utility of a future time step k after some

predicted interaction $ax_{t:k-1}$ under policy π with respect to an environment μ is:

$$V_t^{\pi\mu}(ax_{<k}) := \mathbb{E}_\mu^\pi\left[\sum_{i=k}^{\infty} d(t,i)u(t,ax_{1:i}) \,\middle|\, ax_{<k}\right] \tag{3.5}$$

Specially, for the deterministic policy $\pi\colon (\mathcal{A}\times\mathcal{X})^*\times\mathcal{A}\to\{0,1\}$ or $\pi\colon(\mathcal{A}\times\mathcal{X})^*\to\mathcal{A}$

$$V_t^{\pi\mu}(ax_{<k}) := \mathbb{E}_\mu\left[\sum_{i=k}^{\infty} d(t,i)u(t,ax_{1:i}) \,\middle|\, ax_{<k}\right] \tag{3.6}$$

where $a_i := \pi(ax_{<i})$ for $i \geq k$.
The quantity $V_t^{\pi\mu}(ax_{<k}a_k)$ is defined similarly, except that a_k is now no longer defined by π.

This is essentially a discrete time form of the Bellman equation commonly used in control theory, finance, reinforcement learning and other fields concerned with optimising dynamic systems.

$$\begin{aligned}
&V_t^{\pi\mu}(ax_{<k})\\
&= \mathbb{E}_\mu^\pi\left[\sum_{i=k}^{\infty} d(t,i)u(t,ax_{1:i}) \,\middle|\, ax_{<k}\right]\\
&= \lim_{m\to\infty}\sum_{ax_{k:m}}\left(\sum_{i=k}^{m} d(t,i)u(t,ax_{1:i})\right)\mu^\pi(ax_{k:m}|ax_{<k})\\
&= \sum_{ax_k}\left\{\lim_{m\to\infty}\sum_{ax_{k+1:m}}\left(\sum_{i=k}^{m} d(t,i)u(t,ax_{1:i})\right)\mu^\pi(ax_{k+1:m}|ax_{1:k})\right\}\mu^\pi(ax_k|ax_{<k})\\
&= \sum_{ax_k}\left\{d(t,k)u(t,ax_{1:k}) + \lim_{m\to\infty}\sum_{ax_{k+1:m}}\left(\sum_{i=k+1}^{m} d(t,i)u(t,ax_{1:i})\right)\mu^\pi(ax_{k+1:m}|ax_{1:k})\right\}\mu^\pi(ax_k|ax_{<k})\\
&= \sum_{ax_k}\left(d(t,k)u(t,ax_{1:k}) + V_t^{\pi\mu}(ax_{1:k})\right)\mu^\pi(ax_k|ax_{<k})\\
&= \sum_{ax_{k:m}}\left(\sum_{i=k}^{m} d(t,i)u(t,ax_{1:i}) + V_t^{\pi\mu}(ax_{1:m})\right)\mu^\pi(ax_{k:m}|ax_{<k})
\end{aligned}$$

After possible future history $ax_{<k}$, the *ex post optimal policy* π_t^μ at time t for environment μ is the policy that maximises the ex post expected future utility.

$$\pi_t^\mu := \arg\max_\pi V_t^{\pi\mu}(ax_{<t})$$

It is easy to see, $\pi_t^\mu(a_k|ax_{<k}) = 1$ for the ex post expected future utility maximising action, and zero otherwise, if it exists.

At time t, the maximal achievable expected future utility with possible future history $ax_{<k}$ in environment μ is

$$V_t^{\pi_t^\mu\mu}(ax_{<k})$$

$$
\begin{aligned}
&= \sum_{a_k}\sum_{x_k}\Big(d(t,k)u(t,ax_{1:k}) + V_t^{\pi_t^\mu \mu}(ax_{1:k})\Big)\pi_t^\mu(a_k|ax_{<k})\mu(x_k|ax_{<k}a_k)\\
&= \max_{a_k}\sum_{x_k}\Big(d(t,k)u(t,ax_{1:k}) + V_t^{\pi_t^\mu \mu}(ax_{1:k})\Big)\mu(x_k|ax_{<k}a_k) \qquad \text{(recursive)}\\
&= \max_{a_k}\sum_{x_k}\mu(x_k|ax_{<k}a_k)\Big\{d(t,k)u(t,ax_{1:k})\\
&\qquad + \max_{a_{k+1}}\sum_{x_{k+1}}\mu(x_{k+1}|ax_{1:k}a_{k+1})\Big(d(t,k+1)u(t,ax_{1:k+1}) + V_t^{\pi_t^\mu \mu}(ax_{1:k+1})\Big)\Big\}\\
&= \max_{a_k}\sum_{x_k}\mu(x_k|ax_{<k}a_k)\Big\{\Big(\max_{a_{k+1}}\sum_{x_{k+1}}\mu(x_{k+1}|ax_{1:k}a_{k+1})\Big)\cdot d(t,k)u(t,ax_{1:k})\\
&\qquad + \max_{a_{k+1}}\sum_{x_{k+1}}\mu(x_{k+1}|ax_{1:k}a_{k+1})\Big(d(t,k+1)u(t,ax_{1:k+1}) + V_t^{\pi_t^\mu \mu}(ax_{1:k+1})\Big)\Big\}\\
&= \max_{a_k}\sum_{x_k}\mu(x_k|ax_{<k}a_k)\Big\{\max_{a_{k+1}}\sum_{x_{k+1}}\mu(x_{k+1}|ax_{1:k}a_{k+1})d(t,k)u(t,ax_{1:k})\\
&\qquad + \max_{a_{k+1}}\sum_{x_{k+1}}\mu(x_{k+1}|ax_{1:k}a_{k+1})\Big(d(t,k+1)u(t,ax_{1:k+1}) + V_t^{\pi_t^\mu \mu}(ax_{1:k+1})\Big)\Big\}\\
&= \max_{a_k}\sum_{x_k}\mu(x_k|ax_{<k}a_k)\cdots\max_{a_m}\sum_{x_m}\mu(x_m|ax_{<m}a_m)\Big(\sum_{i=k}^m d(t,i)u(t,ax_{1:i}) + V_t^{\pi_t^\mu \mu}(ax_{1:m})\Big)\\
&= \max_{a_k}\sum_{x_k}\cdots\max_{a_m}\sum_{x_m}\Big(\sum_{i=k}^m d(t,i)u(t,ax_{1:i}) + V_t^{\pi_t^\mu \mu}(ax_{1:m})\Big)\mu(x_{k:m}|ax_{<k}a_{k:m})\\
&= \max_{a_k}\sum_{x_k}\cdots\max_{a_m}\sum_{x_m}\Big(\sum_{i=k}^m d(t,i)u(t,ax_{1:i})\Big)\mu(x_{k:m}|ax_{<k}a_{k:m})\\
&\quad + \max_{a_k}\sum_{x_k}\cdots\max_{a_m}\sum_{x_m}V_t^{\pi_t^\mu \mu}(ax_{1:m})\mu(x_{k:m}|ax_{<k}a_{k:m})\\
&= \lim_{m\to\infty}\max_{a_k}\sum_{x_k}\cdots\max_{a_m}\sum_{x_m}\Big(\sum_{i=k}^m d(t,i)u(t,ax_{1:i})\Big)\mu(x_{k:m}|ax_{<k}a_{k:m}) \qquad \text{(iterative)}
\end{aligned}
$$

The last "=" follows from

$$
\begin{aligned}
\Gamma_{tm} < \infty \implies & V_t^{\pi_t^\mu \mu}(ax_{1:m}) \xrightarrow{m\to\infty} 0\\
\implies & \lim_{m\to\infty}\max_{a_k}\sum_{x_k}\cdots\max_{a_m}\sum_{x_m}V_t^{\pi_t^\mu \mu}(ax_{1:m})\mu(x_{k:m}|ax_{<k}a_{k:m}) = 0
\end{aligned}
$$

$V_t^{\pi_t^\mu \mu}(ax_{<k})$ is obtained by averaging over possible perceptions x_i and by maximizing over the possible actions a_i. This has to be done in chronological order $a_1x_1a_2x_2\ldots$ to correctly incorporate the dependencies of x_i and a_i on the history. For convenience, we will often refer to the Equation (iterative) as the *ExpectiMax operation*, which is similar to the well-known *minimax* strategy in game theory.

It can be seen that, the only use of $\Gamma_{tk} < \infty$ is to guarantee

$$\lim_{m\to\infty}\max_{a_k}\sum_{x_k}\cdots\max_{a_m}\sum_{x_m}V_t^{\pi_t^\mu \mu}(ax_{1:m})\mu(x_{k:m}|ax_{<k}a_{k:m}) = 0$$

So we can drop the summable assumption of $d(t,k)$, and make another assumption instead.

Assumption 1 (**Convergence Condition).**

$$\forall \pi \forall t : \lim_{m\to\infty} \sum_{h_{<m}} V_t^{\pi\mu}(h_{<m})_\mu^\pi(h_{<m}) = 0$$

In general, we have

$$\begin{aligned} V_t^{\pi\mu}(ax_{<k}) &:= \lim_{m\to\infty} \sum_{ax_{k:m}} \left[\sum_{i=k}^{m} d(t,i)u(t,ax_{1:i})\right]_\mu^\pi(ax_{k:m}|ax_{<k}) \\ &= \lim_{m\to\infty} \sum_{i=k}^{m} d(t,i) \sum_{ax_{k:i}} u(t,ax_{1:i})_\mu^\pi(ax_{k:i}|ax_{<k}) \\ &= \lim_{m\to\infty} \sum_{i=k}^{m} d(t,i) R^\pi(ax_{<k})_i \\ &= \begin{bmatrix} d(t,k) & d(t,k+1) & \dots & d(t,\infty) \end{bmatrix} \cdot \begin{bmatrix} R^\pi(ax_{<k})_k \\ R^\pi(ax_{<k})_{k+1} \\ \vdots \\ R^\pi(ax_{<k})_\infty \end{bmatrix} \end{aligned}$$

where

$$R^\pi(ax_{<k})_i := \sum_{ax_{k:i}} u(t,ax_{1:i})_\mu^\pi(ax_{k:i}|ax_{<k})$$

Since

$$V_t^{\pi\mu}(ax_{<k}a_k) = \sum_{x_k} \left(d(t,k)u(t,ax_{1:k}) + V_t^{\pi\mu}(ax_{1:k})\right)\mu(x_k|ax_{<k}a_k)$$

Obviously,

$$\max_{\substack{a_k\in\mathcal{A} \\ \pi:\pi(a_k|ax_{<k})=1}} V_t^{\pi\mu}(ax_{<k}a_k) = \max_\pi V_t^{\pi\mu}(ax_{<k})$$

At time t, the *optimal action* $a_k^{*\mu}$ in the k^{th} $(k \geq t)$ cycle is related to the *ExpectiMax* operation by

$$\begin{aligned} a_k^{*\mu} &:= \arg\max_{a_k} V_t^{\pi\mu}(ax_{<k}a_k) \\ &= a_k^{\pi_t^\mu} \\ &= \arg\max_{a_k} \lim_{m\to\infty} \sum_{x_k} \max_{a_{k+1}} \sum_{x_{k+1}} \cdots \max_{a_m} \sum_{x_m} \left[\sum_{i=k}^{m} d(t,i)u(t,ax_{1:i})\right] \mu(x_{k:m}|ax_{<k}a_{k:m}) \end{aligned}$$

Remark: If μ is allowed to be a semimeasure, the term (recursive) may not be equivalent to the term (iterative), since we only have $\sum_{x_i} \mu(x_i|ax_{<i}a_i) \leq 1$. To rescue the "=", the semimeasure should be *normalized* to a measure. There are several ways to do this, for example

$$\mu'(\epsilon) := 1$$

$$\mu'(x|h) := \frac{\mu(x|h)}{\sum_{x \in \mathcal{X}} \mu(x|h)}$$

or

$$\mu'(x|h) := \begin{cases} \mu(x|h) & \text{if } x \neq 0 \\ 1 - \sum_{x \neq 0} \mu(x|h) & \text{otherwise} \end{cases}$$

The problem is: does π_t^μ exist?

Existence of The Ex Post Equilibrium of Γ_2

Theorem 2 (**Existence of Ex Post Equilibrium of Γ_2**).
Under the assumption1, π_t^μ exists.

To sum up, we have the definition of agent AIμ in *recursive/iterative form.*

Definition 7 (**Agent AIμ in *Recursive/Iterative Form***).

$$V_t^{*\mu}(ax_{<k}) := \max_{a_k \in \mathcal{A}} V_t^{*\mu}(ax_{<k}a_k)$$

$$V_t^{*\mu}(ax_{<k}a_k) := \sum_{x_k \in \mathcal{X}} \mu(x_k|ax_{<k}a_k)\Big(d(t,k)u(t,ax_{1:k}) + V_t^{*\mu}(ax_{1:k})\Big)$$

$$a_k^{*\mu} := \arg\max_{a_k \in \mathcal{A}} V_t^{*\mu}(ax_{<k}a_k)$$

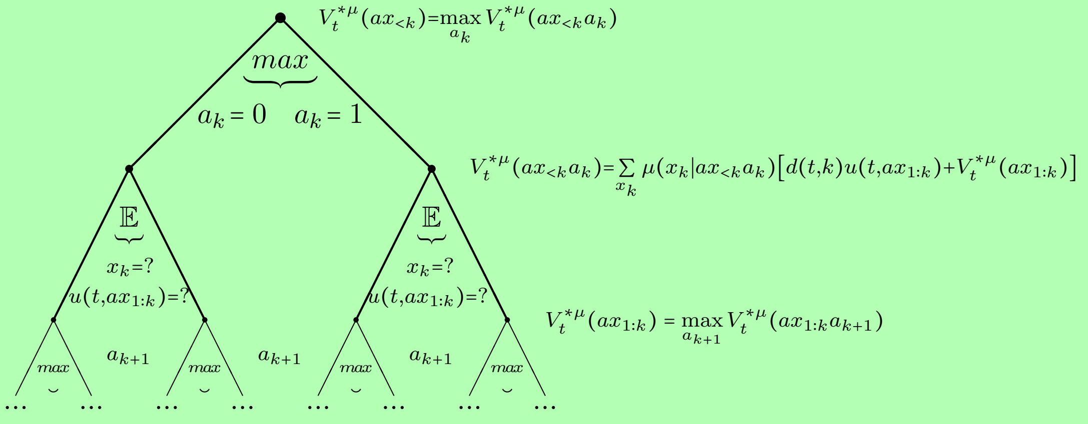


Figure 1: *E*xpectiMax

or[1] *in one line*

[1]If we define

$$V_t^{\pi\mu}(ax_{1:k}) := \mathbb{E}_\mu^\pi\left[\sum_{i=k}^{\infty} d(t,i)u(t,ax_{1:i}) \,\middle|\, ax_{1:k}\right]$$

$$a_k^{*\mu} := \arg\max_{a_k} \lim_{m\to\infty} \sum_{x_k} \max_{a_{k+1}} \sum_{x_{k+1}} \cdots \max_{a_m} \sum_{x_m} \left[\sum_{i=k}^{m} d(t,i)u(t,ax_{1:i})\right] \mu(x_{k:m}|ax_{<k}a_{k:m})$$

§3.1.3 Equivalence of the Functional and Recursive/Iterative AIμ

Theorem 3 (Equivalence Theorem).
The actions of the functional Agent(Definition 4) coincide with the actions of the recursive/iterative Agent(Definition 7) with environments identified by

$$\mu(x(h)|a(h)) = \sum_{q\in\mathcal{Q}} \mu(q)[\![x(h) = q(a(h))]\!] \qquad \text{(subjective-objective)}$$

Proof.

$$\begin{aligned}
&a_k^{*\mu}\\
&= \arg\max_{a_k} \max_{p:p(x_{<k})=a_{<k}a_k} \frac{\sum_{q:q(a_{<k})=x_{<k}} \left[\mu(q)\cdot \sum_{i=k}^{\infty} d(t,i)u(t,ax_{1:i}^{pq})\right]}{\sum_{q:q(a_{<k})=x_{<k}} \mu(q)}\\
&= \arg\max_{a_k} \lim_{m\to\infty} \max_{p:p(x_{<k})=a_{<k}a_k} \frac{\sum_{q:q(a_{<k})=x_{<k}} \left[\mu(q)\cdot \sum_{i=k}^{m} d(t,i)u(t,ax_{1:i}^{pq})\right]}{\sum_{q:q(a_{<k})=x_{<k}} \mu(q)}\\
&= \arg\max_{a_k} \lim_{m\to\infty} \max_{p:p(x_{<k})=a_{<k}a_k} \frac{\sum_{x_{k:m}} \sum_{q:q(a_{1:k}a_{k+1}(x_{1:k})\cdots a_m(x_{<m}))=x_{1:m}} \left[\mu(q)\cdot \sum_{i=k}^{m} d(t,i)u(t,ax_{1:i}^{pq})\right]}{\sum_{q:q(a_{<k})=x_{<k}} \mu(q)}\\
&= \arg\max_{a_k} \lim_{m\to\infty} \max_{p:p(x_{<k})=a_{<k}a_k} \frac{\sum_{x_{k:m}} \left[\sum_{i=k}^{m} d(t,i)u(t,ax_{1:i}^{p})\right] \sum_{q:q(a_{1:k}a_{k+1}(x_{1:k})\cdots a_m(x_{<m}))=x_{1:m}} \mu(q)}{\sum_{q:q(a_{<k})=x_{<k}} \mu(q)}\\
&= \arg\max_{a_k} \lim_{m\to\infty} \max_{p:p(x_{<k})=a_{<k}a_k} \frac{\sum_{x_{k:m}} \left[\sum_{i=k}^{m} d(t,i)u(t,ax_{1:i}^{p})\right] \mu(x_{1:m}|a_{1:k}a_{k+1}(x_{1:k})\cdots a_m(x_{<m}))}{\mu(x_{<k}|a_{<k})}
\end{aligned}$$

---

then we can define AIμ in this way:

$$\begin{aligned}
V_t^{*\mu}(ax_{1:k}) &:= d(t,k)u(t,ax_{1:k}) + \max_{a_{k+1}\in\mathcal{A}} V_t^{*\mu}(ax_{1:k}a_{k+1})\\
V_t^{*\mu}(ax_{1:k}a_{k+1}) &:= \sum_{x_{k+1}\in\mathcal{X}} \mu(x_{k+1}|ax_{1:k}a_{k+1})V_t^{*\mu}(ax_{1:k+1})\\
a_{k+1}^{*\mu} &:= \arg\max_{a_{k+1}\in\mathcal{A}} V_t^{*\mu}(ax_{1:k}a_{k+1})
\end{aligned}$$

$$= \arg\max_{a_k} \lim_{m\to\infty} \max_{p:p(x_{<k})=a_{<k}a_k} \sum_{x_{k:m}} \left[\sum_{i=k}^{m} d(t,i)u(t,ax^p_{1:i})\right] \mu(x_{k:m}|ax_{1:k}a_{k+1}(x_{1:k})\cdots a_m(x_{<m}))$$

$$= \arg\max_{a_k} \lim_{m\to\infty} \max_{a_{k+1}(x_{1:k})} \sum_{x_k} \max_{p:p(x_{1:k})=a_{1:k}a_{k+1}} \sum_{x_{k+1:m}} \left[\sum_{i=k}^{m} d(t,i)u(t,ax^p_{1:i})\right] \mu(x_{k:m}|ax_{1:k}a_{k+1}(x_{1:k})\cdots a_m(x_{<m}))$$

$$= \arg\max_{a_k} \lim_{m\to\infty} \sum_{x_k} \max_{a_{k+1}} \max_{p:p(x_{1:k})=a_{1:k}a_{k+1}} \sum_{x_{k+1:m}} \left[\sum_{i=k}^{m} d(t,i)u(t,ax^p_{1:i})\right] \mu(x_{k:m}|ax_{1:k+1}a_{k+2}(x_{1:k+1})\cdots a_m(x_{<m}))$$

$$= \arg\max_{a_k} \lim_{m\to\infty} \sum_{x_k} \max_{a_{k+1}} \sum_{x_{k+1}} \cdots \max_{a_m} \sum_{x_m} \left[\sum_{i=k}^{m} d(t,i)u(t,ax_{1:i})\right] \mu(x_{k:m}|ax_{<k}a_{k:m})$$

We write $a_i(x_{<i})$ to denote that $a_i(x_{<i})$ is a function of $x_{<i}$, since a_i is uniquely determined by $p(x_{<i}) = a_{<i}a_i$. □

Remark: Given the (subjective) belief system μ in Γ_1 or the true (objective) type of the environment μ in Γ_2, and if equation subjective-objective holds, the behavioural strategy ρ_a in the perfect bayesian equilibrium (ρ,μ) of Γ_1 and the behavioural strategy π in the ex post equilibrium of Γ_2 coincide.

Harsanyi Transformation Again If the agent has the true subjective prior belief μ of the deterministic environment q, the problem of maximizing expected utility is hence being formally solved. However, the problem is: *what if the true prior belief μ of the deterministic environment q in Game Γ_1 or the true type of the environment μ in Game Γ_2 is unknown?* Harsanyi transformation again! This time we resort to the prior belief w of the stochastic environment μ (or the second order prior belief—the prior belief of the prior belied of q). In this way, Game Γ_2 works well. But how to achieve the perfect bayesian equilibrium of Γ_2?

Higher order prior belief is unnecessary, because they can be reduced to sort of second order prior belief (with different initial prior belief w'_ϵ).

$$\begin{aligned}
&\sum_{\rho_n} \rho_{n+1}(\rho_n) \sum_{\rho_{n-1}} \rho_n(\rho_{n-1}) \cdots \sum_{\rho_0} \rho_1(\rho_0)\rho_0(\cdot) \\
&= \sum_{\rho_{n-1}} \sum_{\rho_n} \rho_{n+1}(\rho_n)\rho_n(\rho_{n-1}) \cdots \sum_{\rho_0} \rho_1(\rho_0)\rho_0(\cdot) \\
&= \sum_{\rho_{n-1}} \xi_n(\rho_{n-1}) \cdots \sum_{\rho_0} \rho_1(\rho_0)\rho_0(\cdot) \\
&= \sum_{\rho_0} \xi_1(\rho_0)\rho_0(\cdot) \\
&= \xi_0(\cdot)
\end{aligned}$$

§3.2 Agents in Unknown Environment

§3.2.1 From Bayesian Game Γ_2 to AIξ

Inspired by the subjective-objective equation $\mu(x(h)|a(h)) = \sum_{q\in Q}\mu(q)[\![x(h) = q(a(h))]\!]$, we can define a counterpart of it in Game Γ_2, which induces the subjective prior belief of the stochastic behavior of environment—the mixture environment.

Definition 8 (Subjective Prior Belief/Mixture Environment). *Given a countable environment model class $\mathcal{M}$ and a prior belief $w^\rho_\epsilon > 0$ for each $\rho \in \mathcal{M}$ such that $\sum_{\rho\in\mathcal{M}} w^\rho_\epsilon = 1$, the subjective prior belief/mixture environment is*

$$\xi(x_{1:n}|a_{1:n}) := \sum_{\rho\in\mathcal{M}} w^\rho_\epsilon \rho(x_{1:n}|a_{1:n})$$

Proposition 1. *A subjective prior belief/mixture environment is an environment model.*

From Subjective Belief to Mixture Environment Since we are assuming that the agent does not initially exactly know the true environment, we desire subjective models whose predictive performance improves as the agent gains experience.

The belief system $\mu_e(h)(\theta)$ of Γ_2 should be used to update belief:

$$w^\rho_{ax_{<n}} := \frac{w^\rho_\epsilon \rho(x_{<n}|a_{<n})}{\sum_{\nu\in\mathcal{M}} w^\nu_\epsilon \nu(x_{<n}|a_{<n})} \tag{3.7}$$

It be taken as the posterior belief $w^\rho_{ax_{<n}}$ for environment model ρ.

Given experience $ax_{<n}$, the subjective belief of possible future history $ax_{1:m}$ is

$$\sum_{\rho\in\mathcal{M}} w^\rho_{ax_{<n}} \rho(x_{n:m}|ax_{<n}a_{n:m}) = \frac{\sum_{\rho\in\mathcal{M}} w^\rho_\epsilon \rho(x_{1:m}|a_{1:m})}{\sum_{\nu\in\mathcal{M}} w^\nu_\epsilon \nu(x_{<n}|a_{<n})} = \frac{\xi(x_{1:m}|a_{1:m})}{\xi(x_{<n}|a_{<n})} = \xi(x_{n:m}|ax_{<n}a_{n:m}) \tag{3.8}$$

So, to predict with the convex combination of environment models—with each model weighted by its posterior belief—is equivalent to predict with the mixture environment model ξ.

Ex Interim Expected Utility At time t, with type π, the player a's ex interim expected utility with possible future history h is

$$\begin{aligned}\Upsilon_h(\pi) &= \sum_{\mu\in\mathcal{M}} \mathbb{E}^\pi_\mu\left[u_a|_h(\cdot,(\pi,\mu)) \,\middle|\, h\right]\mu_e(h)(\mu)\\ &= \sum_{\mu\in\mathcal{M}} w^\mu_h V^{\pi\mu}_t(h)\end{aligned}$$

$$= \sum_{\mu \in \mathcal{M}} \frac{\mu(x(h)|a(h)) w_\epsilon^\mu}{\sum_{\nu \in \mathcal{M}} \nu(x(h)|a(h)) w_\epsilon^\nu} \mathbb{E}_\mu^\pi \left[\sum_{i=k}^{\infty} d(t,i) u(t, hh'_{1:i}) \,\middle|\, h \right]$$

Theorem 4 (**Ex Interim Expected Utility as Intelligence Measure**).

$$\Upsilon_{ax_{<k}}(\pi) = V_t^{\pi\xi}(ax_{<k})$$

Proof.

$$\begin{aligned}
&\Upsilon_{ax_{<k}}(\pi) \\
&= \sum_{\mu \in \mathcal{M}} w_{ax_{<k}}^\mu V_t^{\pi\mu}(ax_{<k}) \\
&= \sum_{\mu \in \mathcal{M}} \frac{w_\epsilon^\mu \mu(x_{<k}|a_{<k})}{\sum_{\nu \in \mathcal{M}} w_\epsilon^\mu \nu(x_{<k}|a_{<k})} \mathbb{E}_\mu^\pi \left[\sum_{i=k}^{\infty} d(t,i) u(t, ax_{1:i}) \,\middle|\, ax_{<k} \right] \\
&= \sum_{\mu \in \mathcal{M}} \frac{w_\epsilon^\mu \mu(x_{<k}|a_{<k})}{\sum_{\nu \in \mathcal{M}} w_\epsilon^\mu \nu(x_{<k}|a_{<k})} \lim_{m \to \infty} \sum_{ax_{k:m}} \left(\sum_{i=k}^{m} d(t,i) u(t, ax_{1:i}) \right) {}_\mu^\pi(ax_{k:m}|ax_{<k}) \\
&= \lim_{m \to \infty} \sum_{ax_{k:m}} \left(\sum_{i=k}^{m} d(t,i) u(t, ax_{1:i}) \right) \cdot \sum_{\mu \in \mathcal{M}} \frac{w_\epsilon^\mu \mu(x_{<k}|a_{<k})}{\sum_{\nu \in \mathcal{M}} w_\epsilon^\mu \nu(x_{<k}|a_{<k})} {}_\mu^\pi(ax_{k:m}|ax_{<k}) \\
&= \lim_{m \to \infty} \sum_{ax_{k:m}} \left(\sum_{i=k}^{m} d(t,i) u(t, ax_{1:i}) \right) \cdot \\
&\qquad \sum_{\mu \in \mathcal{M}} \frac{w_\epsilon^\mu \mu(x_{<k}|a_{<k})}{\sum_{\nu \in \mathcal{M}} w_\epsilon^\mu \nu(x_{<k}|a_{<k})} \pi(a_k|ax_{<k}) \mu(x_k|ax_{<k}a_k) \cdots \pi(a_m|ax_{<m}) \mu(x_m|ax_{<m}a_m) \\
&= \lim_{m \to \infty} \sum_{ax_{k:m}} \left(\sum_{i=k}^{m} d(t,i) u(t, ax_{1:i}) \right) \cdot \sum_{\mu \in \mathcal{M}} \frac{w_\epsilon^\mu \mu(x_{<k}|a_{<k})}{\sum_{\nu \in \mathcal{M}} w_\epsilon^\mu \nu(x_{<k}|a_{<k})} \prod_{i=k}^{m} \pi(a_i|ax_{<i}) \mu(x_{k:m}|ax_{<k}a_{k:m}) \\
&= \lim_{m \to \infty} \sum_{ax_{k:m}} \left(\sum_{i=k}^{m} d(t,i) u(t, ax_{1:i}) \right) \cdot \prod_{i=k}^{m} \pi(a_i|ax_{<i}) \sum_{\mu \in \mathcal{M}} \frac{w_\epsilon^\mu \mu(x_{1:m}|a_{1:m})}{\sum_{\nu \in \mathcal{M}} w_\epsilon^\mu \nu(x_{<k}|a_{<k})} \\
&= \lim_{m \to \infty} \sum_{ax_{k:m}} \left(\sum_{i=k}^{m} d(t,i) u(t, ax_{1:i}) \right) \cdot \prod_{i=k}^{m} \pi(a_i|ax_{<i}) \xi(x_{k:m}|ax_{<k}a_{k:m}) \\
&= \lim_{m \to \infty} \sum_{ax_{k:m}} \left(\sum_{i=k}^{m} d(t,i) u(t, ax_{1:i}) \right) \cdot {}_\xi^\pi(x_{k:m}|ax_{<k}a_{k:m}) \\
&= \mathbb{E}_\xi^\pi \left[\sum_{i=k}^{\infty} d(t,i) u(t, ax_{1:i}) \,\middle|\, ax_{<k} \right] \\
&= V_t^{\pi\xi}(ax_{<k})
\end{aligned}$$

□

To achieve the perfect bayesian equilibrium $((\pi, \mu), w)$, we have to chose

$$\pi_t^\xi := \arg\max_\pi \Upsilon_{ax_{<t}}(\pi) = \arg\max_\pi V_t^{\pi\xi}(ax_{<t})$$

It is easy to see, $\pi_t^\xi(a_t|ax_{<t}) = 1$ for the ex interim expected future utility maximising action, and zero otherwise.

π_t^ξ depends on t and is used only in step t to determine a_t by $\pi_t^\xi(x_{<t}|a_{<t}) = a_{<t}a_t$.

Similar to Definition4, we can define agent AIξ in ***functional form*** as follows.

Definition 9 (**Agent AIξ in *Functional Form***)**.**

$$V_t^{\pi\mu}(ax_{<k}) := \mathbb{E}_\mu^\pi\left[\sum_{i=k}^{\infty} d(t,i)u(t,ax_{1:i}) \,\middle|\, ax_{<k}\right] \quad \text{(Ex Post)}$$

$$V_t^{\pi\xi}(ax_{<k}) := \frac{\sum_{\mu\in\mathcal{M}} w_\epsilon^\mu \mu(x_{<k}|a_{<k}) \cdot V_t^{\pi\mu}(ax_{<k})}{\sum_{\nu\in\mathcal{M}} w_\epsilon^\nu \nu(x_{<k}|a_{<k})} \quad \text{(Ex Interim)}$$

$$\pi_t^\xi := \arg\max_\pi V_t^{\pi\xi}(ax_{<t})$$

$$\pi^\xi(x_{<t}) := \pi_t^\xi(x_{<t}|\pi_{t-1}^\xi(x_{<t-1}|\ldots\pi_1^\xi(\epsilon)\ldots)) \quad \text{(Perfect Bayesian)}$$

or in one line

$$a_k^{*\xi} := \arg\max_{a_k\in supp(\pi(\cdot|ax_{<k}))} \max_\pi \frac{\sum_{\mu\in\mathcal{M}} w_\epsilon^\mu \mu(x_{<k}|a_{<k}) \mathbb{E}_\mu^\pi\left[\sum_{i=k}^{\infty} d(t,i)u(t,ax_{1:i}) \,\middle|\, ax_{<k}\right]}{\sum_{\nu\in\mathcal{M}} w_\epsilon^\nu \nu(x_{<k}|a_{<k})}$$

Or in another way, just replace μ with ξ in Definition7, we get agent AIξ in recursive/iterative form.

	Game Γ_1	Game Γ_2
Ex Post Equilibrium	Deterministic	AIμ(recursive/iterative)
Ex Ante Equilibrium	AIμ(functional)	AIξ (functional)

Table 1: Intelligence is an equilibrium, we just have to identify the game.

§3.2.2 Theoretical Properties

Theorem 5.

$$\max_\pi \Upsilon_h(\pi) = \Upsilon_h(\pi_t^\xi)$$

We now show that if there is a good model of the (unknown) environment in $\mathcal{M}$, an agent using the mixture environment model $\xi(x_{1:n}|a_{1:n})$ will predict well.

Definition 10 (**Pareto-Optimality**).
For any t and h, a policy $\tilde{\pi}$ is called Pareto-optimal if

$$\nexists\pi:\ \forall h \forall\rho \in \mathcal{M}\left[\left(V_t^{\pi\rho}(h) \geq V_t^{\tilde{\pi}\rho}(h)\right) \ \&\ \exists\mu \in \mathcal{M}\left(V_t^{\pi\mu}(h) > V_t^{\tilde{\pi}\mu}(h)\right)\right]$$

Theorem 6 (**Pareto-Optimality**).
π_t^{ξ} *is Pareto-optimal.*

Theorem 7 (**Convergence Theorem**).
Let μ be the true environment. The μ-expected squared difference of μ and ξ is bounded as follows. For all $n \in \mathbb{N}$, for all $a_{1:n}$,

$$\sum_{k=1}^{n}\sum_{x_{1:k}}\mu(x_{<k}|a_{<k})\Big(\mu(x_k|ax_{<k}a_k) - \xi(x_k|ax_{<k}a_k)\Big)^2 \leq \min_{\rho\in\mathcal{M}}\left\{-\ln w_\epsilon^\rho + D_{1:n}(\mu\|\rho)\right\}$$

where

$$D_{1:n}(\mu\|\rho) := \sum_{x_{1:n}}\mu(x_{1:n}|a_{1:n})\ln\frac{\mu(x_{1:n}|a_{1:n})}{\rho(x_{1:n}|a_{1:n})}$$

In Theorem7, take the supremum over n in the r.h.s and then the limit $n \to \infty$ on the l.h.s. If $\sup_n D_{1:n}(\mu\|\rho) < \infty$ for the minimising ρ, the infinite sum on the l.h.s can only be finite if $\xi(x_k|ax_{<k}a_k)$ converges sufficiently fast to $\mu(x_k|ax_{<k}a_k)$ for $k \to \infty$ with probability 1, hence ξ predicts μ with rapid convergence.

Theorem 8 (**Convergence of Universal to True Value**).
If the history $ax_{<k}$ is generated by policy π (and environment μ), then

$$\mathbb{E}_\mu^\pi\left[\left(V_t^{\pi\xi}(ax_{<k}) - V_t^{\pi\mu}(ax_{<k})\right)^2 \,\middle|\, ax_{<k}\right] \leq 2\left(\max_h u(t,h)\right)^2(\Gamma_{tk})^2\left(D_{1:\infty} - D_{<k}\right)$$

Definition 11 (Time-Consistency).

$$\forall\mu\forall h\forall j,k \leq |h|:\ \pi_j^{*\mu}(h) = \pi_k^{*\mu}(h)$$

For example, if

$$\forall t,k: d(t,k)u(t,h_{1:k}) = d(k,k)u(k,h_{1:k})$$

let

$$u(h_{1:k}) := d(k,k)u(k,h_{1:k})$$

then

$$V^{*\mu}(ax_{<k}) = \lim_{m\to\infty}\max_{a_k}\sum_{x_k}\cdots\max_{a_m}\sum_{x_m}\left[\sum_{i=k}^{m}u(ax_{1:i})\right]\mu(x_{k:m}|ax_{<k}a_{k:m})$$

and

$$a_k^{*\mu} = \arg\max_{a_k} \lim_{m\to\infty} \sum_{x_k} \cdots \max_{a_m} \sum_{x_m} \left[\sum_{i=k}^{m} u(ax_{1:i}) \right] \mu(x_{k:m}|ax_{<k}a_{k:m})$$

This is a special case of *Time-Consistency*11.

$\pi^\xi = \pi_t^\xi$ under the assumption of *Time-Consistency*.

§3.3 AIXI

Reinforcement-Learning Agent For a reinforcement-learning agent π_{rf}, $x_t = \langle o_t, r_t \rangle \in \mathcal{X} = \mathcal{O} \times \mathcal{R}$ and $u(h) := u(t,h) = r_{|h|}$ for all t.

$$d(t,k) = \begin{cases} 1 & \text{if } k - t < m_t \\ 0 & \text{otherwise} \end{cases}$$

where m_t is some variable horizon.

Goal-Seeking Agent For a goal-seeking agent π_g, $u(h) = 1$ if the goal is reached at time $|h|$ and is 0 otherwise. The goal can be reached at most once $\sum_{t=0}^{\infty} u(h_t) \leq 1$. Exponential/Geometric Discounting: $d(t,k) = \gamma^{k-t}$ for $0 < \gamma < 1$, or Hyperbolic/Harmonic Discounting: $d(t,k) = \frac{1}{(1+\alpha(k-t))^\beta}$ for $\alpha > 0, \beta > 1$.

Prediction-Seeking Agent For a prediction-seeking agent π_p, $u(h) = 1$ if $x_{|h|+1} = \arg\max_{x\in\mathcal{X}} \mu(x|ha)$ and 0 otherwise. $d(t,k) = 1$ if $k \leq m_t$ and 0 otherwise.

Knowledge-Seeking Agent For a knowledge-seeking agent π_k, we chose

$$u(t, ax_{1:k}) = -\mu(x_{t:k}|ax_{<t}a_{t:k})$$

or

$$u(t, ax_{1:k}) = -\log_2 \mu(x_{t:k}|ax_{<t}a_{t:k})$$

or

$$u(t, ax_{1:k}) = -\log_2 \xi(x_{t:k}|ax_{<t}a_{t:k})$$

or

$$u(t, ax_{1:k}) = \sum_{\nu\in\mathcal{M}} w^\nu_{ax_{1:k}} \log_2 \frac{w^\nu_{ax_{1:k}}}{w^\nu_{ax_{<t}}}$$

$$= \sum_{\nu \in \mathcal{M}} \frac{w_\epsilon^\nu}{\xi(ax_{1:k})} \cdot \left(-\nu(x_{1:k}|a_{1:k}) \log \frac{\nu(x_{t:k}|ax_{<t}a_{t:k})}{\xi(x_{t:k}|ax_{<t}a_{t:k})} \right)$$

$d(t,k) = 1$ if $k = m_t$ and 0 otherwise.

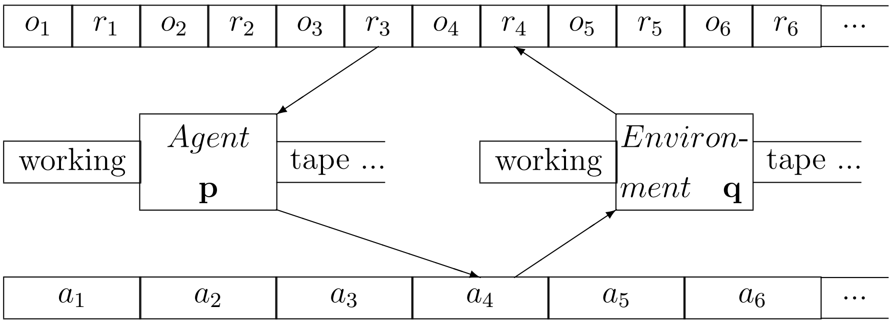
Self-Modifying Agent The self-modifying agent π_{sm} has two parts: its formal description (its code) $c_t \in \mathcal{C}$ and the code executor $\mathcal{E}$ (an Oracle). The set $\mathcal{C}$ contains all programs whose length (in the language of $\mathcal{E}$) is less than a small, arbitrary value. The code executor takes a history h and a program c_{t-1}, and executes the latter to produce an output $a_t = \langle a'_t, c_t \rangle := \mathcal{E}(c_{t-1}, h)$ (with $a_t \in \mathcal{A} = \mathcal{A}' \times \mathcal{C}$) composed of the next action a'_t and new description c_t. Assume the environment has read-access to the agent's code. The initial agent program is:

$$\begin{aligned} c_0(h) = \text{``} \quad & \arg\max_{a \in \mathcal{A}} V_{|h|+1}^{*\mu}(h,a); \\ & V_t^{*\mu}(h, a = \langle a', c \rangle) = \sum_{x \in \mathcal{X}} \mu(x|ha) \Big(d(t,|h'|) u(t,h') + V_t^{*\mu}(h', c(h')) \Big), \\ & h' = hax \text{ ''} \end{aligned}$$

Self-Modifying Agent with Delusion Box Suppose the self-modifying agent π_{smdb} has a delusion box d_t to modify its inputs x'_t (output by the inner environment according to a'_t) to x''_t before they touch its sensors, and the agent's code is fully modifiable, both by the agent itself through c'_t and by the environment, which changes c'_t and returns the new agent's code c_t.

$$\begin{aligned} c_0(h) = \text{``} \quad & \arg\max_{a \in \mathcal{A}} V_{|h|+1}^{*\mu}(h,a); \\ & V_t^{*\mu}(h, a = \langle d, a', c' \rangle) = \sum_{x = \langle x'', c \rangle \in \mathcal{X}} \mu(x|ha) \Big(d(t,|h'|) u(t,h') + V_t^{*\mu}(h', c(h')) \Big), \\ & h' = hax \text{ ''} \end{aligned}$$

AIXI: The Universal Bayesian Agent Theorem7 motivates the construction of Bayesian agents that use rich model classes. The AIXI agent is the limiting case of this viewpoint, by using the largest model class expressible on a Turing machine.



The action picked by AIXI at time t, having executed actions $a_1 a_2 \ldots a_{t-1}$ and having received the sequence of observation-reward pairs $or_1 or_2 \ldots or_{t-1}$ from the environment, is given by:

$$\begin{aligned} a_t^{*\xi_U} &= a_t^{\pi_t^{\xi_U}} \\ &= \arg\max_{a_t} V_t^{\pi_t^{\xi_U}\xi_U}(aor_{<t}a_t) \\ &= \arg\max_{a_t} \sum_{or_t} \ldots \max_{a_{m_t}} \sum_{or_{m_t}} \left[\sum_{i=t}^{m_t} r_i \right] \xi_U(or_{t:m_t} | aor_{<t} a_{t:m_t}) \end{aligned}$$

where

$$\xi_U(or_{1:n}|a_{1:n}) := \sum_{\nu \in \mathcal{M}_C} 2^{-K(\nu)} \nu(or_{1:n}|a_{1:n})$$

is an enumerable chronological semimeasure, and $\nu(or_{1:n}|a_{1:n})$ is the probability of observing $or_1 or_2 \ldots or_n$ given actions $a_1 a_2 \ldots a_n$, class $\mathcal{M}_C$ consists of all *enumerable chronological semimeasures*, which includes all computable ν.

Theorem7 shows for all $n \in \mathbb{N}$ and for all $a_{1:n}$

$$\sum_{k=1}^{n} \sum_{x_{1:k}} \mu(x_{<k}|a_{<k}) \Big(\mu(x_k|ax_{<k}a_k) - \xi_U(x_k|ax_{<k}a_k) \Big)^2 \overset{+}{\leq} K(\mu) \ln 2 \tag{3.9}$$

AIXI: The Reinforcement-Learning Agent AIXI can be seen (Theorem1) as a special case of reinforcement-learning agent: $\mu(q) = 2^{-|q|}$

$$\mu(x(h)|a(h)) = \sum_{q:q(a(h))=x(h)} 2^{-|q|} =: M(x(h)|a(h)) \overset{\times}{=} \xi_U(x(h)|a(h))$$

AIXI in iterative form:

$$a_t^{*M} = \arg\max_{a_t} \sum_{or_t} \ldots \max_{a_{m_t}} \sum_{or_{m_t}} \left[\sum_{i=t}^{m_t} r_i \right] M(or_{t:m_t}|aor_{<t}a_{t:m_t}) \tag{3.10}$$

The equivalence(Theorem3) of agent in the functional and iterative form is true for every chronological semimeasure μ, in particular for M.

AIXI in functional form:

$$a_t^{*M} = \arg\max_{a_t} \max_{p:p(or_{<t})=a_{<t}a_t} \frac{\sum_{q:q(a_{<t})=or_{<t}} \left[2^{-|q|} \cdot \sum_{i=t}^{m_t} r_i\right]}{\sum_{q:q(a_{<t})=or_{<t}} 2^{-|q|}} \tag{3.11}$$

Essentially AIXI is a generalisation of Solomonoff induction to the reinforcement learning setting, that is, where the agent's actions can influence the state of the environment. Conversely, Solomonoff induction can be seen as a special case of AIXI, when the actions taken by the agent have no effect on the environment: $\mathcal{A} = \varnothing$. And when $\mathcal{A} = \varnothing$, the convergence theorem7 is reducible to Solomonoff's completeness theorem.

§3.4 Leibniz vs. UAI

- Ontology—God does not play dice directly with us, but plays dice indirectly through some Universal Turing machine to offer us the freedom to realize any possible regular world. Our world is a collection of programs (Monads) interacting with each other, by computing/simulating each other or themselves.

- God's Perfection(sovereign goodness, omniscience and omnipotence)—God's dice is absolutely fair and God never play tricks; God sets his ultimate utility is constant as long as the output world is regular and he knows it, in other word, God allows us to realize any possible world; God has a universal monotone Turing machine, and he never creates any totally random world, and the way God throw the dice is effective all the time. In other words, God offers us the fullest freedom to chose the most perfect world, that is to say, the one which is at the same time the simplest in hypothesis and the richest in phenomena.

- Time—Our physical notion of a one-dimensional time needs not necessarily be equated with the step-by-step one-dimensional flow of the computation. Computations do not need to evolve the universe, but merely describe it (defining all its relations). The role of the simulating computer is not to compute the history of our universe, but to specify it. Similarly, all of the physical phenomena—space-time, energy, all the corporeal mass—should be explained by the process of computation.

- Free Will—We can choose the best possible world by maximizing the expected future utility, and God never directly tell us whether we have completely grasped the true environment or not. We takes the role of the creator. We are the creators of our own universes.

- Probability—Probability is the uncertainty of possible world, although every possible world can be deterministic. Universal Probability is the limit of the relative frequency of the consistent possible worlds over all possible worlds, and we can use it to predict the true universe.

- Continuum—Within each universe all observable quantities are discrete, but the multiverse as a whole is a continuum. When the equations of quantum theory describe a continuous but not-directly-observable transition between two values of a discrete quantity, what they are telling us is that the transition does not take place entirely within one universe. So perhaps the price of continuous motion is not an infinity of consecutive actions, but an infinity of concurrent actions taking place across the multiverse.

- Intelligence—Intelligence of an 'agent' is an Perfect Bayesian Equilibrium of the Incomplete Information game with observable actions played against imaginary possible worlds ($\mathcal{M}_U$) only if our subjective belief system is the universal prior; or equivalently, Intelligence can be seen as an Ex Post Equilibrium of the Incomplete Information game with observable actions played against imaginary possible worlds ($\mathcal{M}_U$) only if we pretend that the true environment is the universal mixture environment; or equivalently, Intelligence can also be regarded as an Perfect Bayesian Equilibrium of the Incomplete Information game with observable actions played against Leibniz's possible worlds ($\mathcal{M}_D$) only if our subjective belief system is the Solomonoff prior.

In conclusion, we live in a computable universe, which agrees with Wheeler's philosophy "it from bit" and Wolfram's "all is computation". AIXI is the most intelligent creator among us, according to Leibniz's philosophy, AIXI is the greatest monad—the real "God".

References

[1] C. S. Calude. Information and Randomness: An Algorithmic Perspective. *Springer, Berlin, 2nd edition*, 2002.

[2] G. J. Chaitin. Algorithmic Information Theory. *Cambridge University Press, Cambridge*, 1987.

[3] R. G. Downey and D. R. Hirschfeldt. Algorithmic Randomness and Complexity. *Springer-Verlag*, New York, 2010.

[4] Dov M. Gabbay, S. Hartmann and J. Woods. Handbook of the History of Logic. Volume 10: Inductive Logic. *Elsevier*, 2011.

[5] M. Hutter. Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability. *Springer*, 2005.

[6] M. Hutter. On universal prediction and Bayesian confirmation. *Theoretical Computer Science*, 384(1):33–48, 2007.

[7] E. T. Jaynes. Probability Theory: The Logic of Science. *Cambridge University Press, Cambridge, MA*, 2003.

[8] S. Legg. Machine Super Intelligence. *PhD thesis, Department of Informatics, University of Lugano*, 2008.

[9] M. Li and P. M. B. Vitányi. An Introduction to Kolmogorov Complexity and its Applications. *Springer, Berlin, 3rd edition*, 2008.

[10] A. Nies. Computability and Randomness. Volume 51 of Oxford Logic Guides. *Oxford University Press, Oxford*, 2009.

[11] P.G. Odifreddi. Classical Recursion Theory. Volume2. *Elseviser*, 1999.

[12] M. Osborne and A. Rubinstein. A Course in Game Theory. *Cambridge, MIT Press*, 1994.

[13] S. Rathmanner and M. Hutter. A philosophical treatise of universal induction. *Entropy*, 13(6):1076–1136, 2011.

[14] M. Ring and L. Orseau. Delusion, survival, and intelligent agents. *In: AGI 2011*, Springer, Heidelberg.

[15] J. Schmidhuber. Hierarchies of generalized Kolmogorov complexities and nonenumerable universal measures computable in the limit. *International Journal of Foundations of Computer Science*, 2002.

[16] J. Schmidhuber. The speed prior: A new simplicity measure yielding near-optimal computable predictions. In *Proc. 15th Conf. on Computational Learning Theory (COLT'02)*, volume 2375 of *LNAI*, pages 216–228, Sydney, 2002. Springer, Berlin.

[17] J. Schmidhuber. Optimal ordered problem solver. *Machine Learning*, 54: 211–254, 2004.

[18] J. Schmidhuber. Ultimate cognition à la Gödel. *Cognitive Computation*, 2009.

[19] Y. Shoham and K. Leyton-Brown. Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations. *Cambridge University Press*, 2009.

[20] R. J. Solomonoff. A formal theory of inductive inference: Parts 1 and 2. *Information and Control*, 7:1–22 and 224–254, 1964.

[21] R. J. Solomonoff. Complexity-based induction systems: Comparisons and convergence theorems. *IEEE Transactions on Information Theory*, IT-24:422–432, 1978.

[22] R. J. Solomonoff. Progress In Incremental Machine Learning. TR IDSIA-16-03. 2003

[23] M. Tegmark. The Mathematical Universe. *Foundations of Physics*, 38: 101–150, 2008.

[24] J. Veness, K. S. Ng, M. Hutter, W. Uther, and D. Silver. A Monte Carlo AIXI approximation. *Journal of Artificial Intelligence Research*, 40:95–142, 2011.

[25] J. von Neumann. The general and logical theory of automata. *Cerebral mechanisms in behaviour*, 1–41, 1951.

[26] J. A. Wheeler. Information, Physics, Quantum: The Search for Links. *Physics Department, Princeton University*, 1990.

[27] F. Ye. Strict Finitism and the Logic of Mathematical Applications. *Synthese Library, vol. 355, Springer*, 2011.

[28] H. Zenil. (Editor) A Computable Universe: Understanding and Exploring Nature as Computation. *World Scientific Publishing Company*, 2012.