

# On the Intelligibility of (dynamic) Systems and Conceptual Uncertainty

A collection of my thoughts on intelligibility. An attempt to edge towards a basic theory for understanding dynamic systems by computationally bounded observers. While the aim is to have practical implications for the design of sophisticated observers, these ideas are quite far-reaching and do tend to border on philosophy (an inevitability, perhaps).



OrbitMines Research

Fadi Shawki

[shawkifadi@gmail.com](mailto:shawkifadi@gmail.com)



FadiShawki



@\_FadiShawki



Fadi#2854

December, 2022

## Abstract

Central to the theme of this paper lies the conceptual framework of framing all ideas and concepts (or generally; information) as perceived superpositions. Building on prior ideas of an observer's limitations and how such a framework is a necessity. One might believe there is such a thing as objective reality, but perfect knowledge of such a thing seems to lie in the space of impossibilities. I consider this mentioning of 'perfect' and other extremes quite susceptible to misinterpretation, which I believe are necessary conditions and imperfections of observers.

And yet, we know enough of to be practical, we know enough in order to manipulate our environment. How then, should we go about instilling such properties in a system of our making? I'll first try to dissect the problems which come up when we humans try to pin abstract concepts down. While doing so, I'll propose a way to re-conceptualize these into concise practical problems to tackle.

## I. Systems, only Systems

As with many ideas, we associate similar concepts with wildly different words; which is quite useful to disambiguate between domains, but it creates all kinds of difficulties when trying to generalize these ideas. While I'll go into detail about interpreting words and meanings, debating such ontological discussions is not my purpose with this paper. While I deem them useful (and believe they likely push words into an equilibrium of their meaning), I will only mention them to come up with useful collective terms we can use in our framework.

One such a term I'll use is a system (or rather: the perception of one). Abstractly, I'll define this as some arbitrary level of description at which computation happens. So in some sense it's the idea of modelling interactions. To put it into perspective some random examples are: Two 'isolated' particles colliding, a singular logic gate, a computer, a function, a keyboard, the internet, a human, a planet, the universe. Note that this perception of a system happens within a system we could call an (intelligent) observer; systems can be polymorphic, overlapping, approximate, recursive, mutable, ..., and as such their interpretation is non-trivial [1] [2].

Conceivably, the only way to understand systems, or at the very least to 'do useful things', is to interact with them. Note that when I say 'interact', I mean so in a way that observing (by whatever means) falls within that definition. If in any way information/interactions can be transferred between two conceived systems, they will fall under what I'd incorrectly call 'interaction'. Which is quite related to the old debate that is the possibility of a 'non interacting' observer [3]. Others have gone to great length discussing these issues [5]. For our purposes I'd thus wrongly assume that any system always exists within another system [7]. Therefore, a practical view on this is merely: if systems cannot interact with one-and-other, they won't ever know of each-other's existence. But I don't consider this approach as rigorous.

## II. Equivalence & Extremes

A seductive take on equivalence, is that the best (; or only perfect) description of a system is that system itself [8]. Any perception of a system, any description of a system, is necessarily flawed. When we try and describe a way in which something is equivalent, we aim for an approximate seemingly useful solution. This is the pocket in which we can operate on intelligibility [9]. This has all kinds of conceptual consequences which are central to this paper's approach. One of the critical ones being that extremes such as 'proof', 'real' or 'truth' invoke things that are generally unknowable,

and that a claim like this is incredibly self-referential. Following from that is that in a certain light; for our perceived purposes many extremes (; or unknowable things) don't matter to us, unless edging toward them is deemed useful.

As a quick and relevant example of this is the concept of something 'arbitrary'; say some arbitrary sequence.

Sequence: [ A, B, ... ] = AB...

Like many concepts in our languages, the nature of its abstractness allows for the creation of something 'ideally/perfectly/truly' arbitrary. Which following my arguments don't exist in reality other than their perception. It is only arbitrary, conceptually. Some system generated this seemingly arbitrary sequence. 'Arbitrary' like many concepts, hold up only in a certain frame of reference [12]. There is a certain complex ambiguity related to concepts which seems context-dependant.

### III. Ambiguity: Conceptual Superpositions

The possibility of perceived ambiguity demands a solution, in some ways this problem seems the most critical to me. In part because of its closeness to concepts like modelling, (conflicted) types, states of flux, symmetries and superpositions. If one tries to observe some phenomenon which continuously changes its state; a dynamic system, through time a non-trivial superposition arises.

SuperPosition: AB... = A | B | ...

Generalizing these ideas and using the term superposition for this, seems quite intuitive. Quite nicely following from this when one tries to disambiguate (; when one tries to collapse the superposition), one gets a representation which seems more actionable, but is necessarily a simplification [13].

To make sense of the intelligible, strategies for collapsing these conceptual superpositions will be required. Among the obvious candidates being: more information or more compute. The strategy I will spend more time on later is that of conceptual extremes. As an example: for an observing system, 'ignorance' can be incredibly useful for the perception of having disambiguated.

### IV. Conceptual Uncertainty

When we think about information 'which we currently possess', there's a subtle assumption of (at the very least partial) persistence of that information. It seems unlikely that one can do anything without some form of memory. Within that memory some form of perceived uncertainty can be encoded. But abstractly, if we call to some 'higher truth', we could say that any perceived information at any step of perceived computation can be faulty for arbitrary reasons. Or more generally, since our descriptions of systems are inescapably flawed, so is any assumption of persistence of information. Any perceived information holds with it an inherent property of 'certainty seems impossible'.

A = A | Uncertainty

Note that, like unknowable extremes, there's a difference between perceived uncertainty and 'actual (unknowable) uncertainty' [14].

Unfortunately, having such inherent properties, since they are self-referential, leads to infinite regression. And doesn't help with decidability; or practicality. How then, does one approach infinities like uncertainty?

### V. Collapsing Infinities with Extremes

A useful extreme to point out here is 'trust'. If verification of uncertainty results into uncertainty, then trust is the point at which we simply cut off the infinite regression and say: "Let's just trust this output". Important to note here is that this 'trusting' can be as non-trivial as the thing it's being ascribed to. It can be temporary, it can have random 'patrol checks', one can later realize this 'trust' was incorrectly placed, but importantly, it cannot be perfect; it can only be useful [16].

Extremes then, are extremely related to inhibitory behavior. They reduce the perceived space into a smaller space, which is more easily interpreted, but necessarily flawed. This cutoff is a necessity: necessities demand practicality. This introduces us to the concept of priorities (; generically: pressures); what are we optimizing for, what does one spend their time on. As an example I previously mentioned; there's a certain ignorance associated with these extremes.

This example seems incredibly relevant: If we take a neutral position on information processing, we'd say that there is no such thing as noise. When we invoke the concept of noise, we invoke a bias to filter for a specific kind of information we deem relevant. When later deemed that process was 'incorrect' (concluded using similar systems), we could call it ignorant, naive or any number of other things [17]. But what we are doing when filtering out noise is removing information: "There might have been a reason for that noise and it's perhaps worth exploring."

Avoidance of ignorance can be optimized for (; or prioritized), it cannot be exterminated. The perception of being ignorant however, is another story entirely. Note that a similar thing can be said about the alignment of systems' priorities. It like many problems which seem to have reached an infliction point of sufficient complexity; seem generally unsolvable. All those problems then come



down to this general solution; we seem to be unable to generally solve these kinds of problems. All one can do is try to approximate solutions. Following previous arguments, even if non-approximate solutions are found, conceptual uncertainty still persists.

It is also worth pointing out that collapsing these infinities don't necessarily need a definite answer. When we take a look at a simple example of an undecidable problem (given our language heuristics). One can collapse this into a (probabilistic) superposition, which is more useful to us than regressing infinitely.

This sentence is false = T | F

## VI. Irreversibility

When determining equivalence or possible ways to transform the current into the future, a familiar extreme is used: reversibility. If we take language as an example it's the perceived reversible relation between object and name. The name invokes a (partial) reconstruction of an object, exposure to such an object would collapse within the observing system to a name. This simple idea of reversibility permeates everything intelligible. Compilers, interpreters, content-addressing (be it by name or hash), ..., theorem proving. A proper way of describing reversibility, is a certain bi-directional relationship between systems. Following from that is that concepts like reversibility, symmetries and permutation invariance are inhibitory extremes on a critical context: The actual sequence of steps of computation which were performed to reach a certain state.

It is that relation which defines something reversible. A certain dependence on the fact that the sequence of steps to get to the desired proven symmetry is irrelevant. The important part being here, that in the light of conceptual uncertainty, it actually does matter which computations were performed, because it could be influenced by something arbitrarily uncertain or unknown. Even if previously one had determined or proved that they are in fact 'the same thing'. An adaptive system hoping to tackle the space of intelligibility, must be able to adapt to falsified things previously thought to be true, decide to combat the falsification because that must surely be false, or 'simply' take a superposition of the situation to be collapsed at a later time. And so time continues...

## Wrapping up

As it goes with theories, this one in particular is as useful as it is vague. However, it has helped me define with better clarity the monstrosity that is this problem. While I'm currently working on implementations on top of these ideas I didn't feel the need to include them in this summary of thoughts.

These implementations are operating in the world of functional equivalence, compilers and notably as a practical first milestone an attempt at the [Hutter, Marcus. "Hutter Prize"](#). These seem like obvious successors to clear first. I'm hoping to share my progress on this in the coming year of 2023.

While generalizing many ideas I've likely inescapably fallen to the 'having described both low- & high-level concepts with the same names' as Hofstadter described. This was done with a certain intentionality in an attempt to frame many problems as the same generic thing. I intend to point out individualistic differences between specific implementations of concepts where deemed useful.

Below a list of references can be found which I explicitly mentioned in this paper. I intend to release a full list of books, papers and other information I've exposed myself to leading up to this paper for additional reference in the future in the case that I fell short of mentioning them here.

## Footnotes & References

[1] [Hofstadter, Douglas R. "Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought" Basic books \(1995\)](#)

[2] [Hofstadter, Douglas R. "Gödel, escher, bach" New York: Basic books \(1979\)](#)

[3] While reading [Kumar, Manjit. "Quantum: Einstein, Bohr and the great debate about the nature of reality" Icon Books Ltd \(2008\)](#) I realized that this dichotomy between Bohr & Einstein is incredibly relevant, even for 'high-level' systems. To my mind summarized as the things knowable against some [\[5\] Active Reality. "And how incredibly interdependent systems are \(whether fuzzy or not\) hyp. 1961](#)

[7] A possibly unanswerable paradox lies within that assumption, however. Namely that of a paradoxical universe; or more accurately the unknowable properties of its existence. And for

practical purposes of systems that doesn't matter much (except of course in the case of wanting to figure more out about it; if that is deemed useful).

[8] [Davis, Randall, Howard Shrobe, and Peter Szolovits. "What is a knowledge representation?" AI magazine 14.1 \(1993\)](#)

[9] Similar to Wolfram's idea of computational (ir)reducibility: [Wolfram, Stephen, and M. Gad-el-Hak. "A new kind of science?" Appl. Mech. Rev. 56.2 \(2003\)](#) & [Wolfram, Stephen. "A Project to Find the Fundamental Theory of Physics" Wolfram Media, Inc. \(2020\)](#)

[12] Note that in the cases of both for and against determinism the same arguments can be made: Notably that either is generally unknowable to the observer. The same can be said of the concept of infinity: For the observer's purposes, infinity is only so conceptually. It would make little difference to the observer even if infinite things can exist beyond their conception; either perceived infinities keep computing, or they are halted.

[13] This is fairly analogous to mapping a complex visual superposition to a single (numeric & low-cardinality) symbol. Doing so is useful, but hides the fact that such visual mapping is faulty. Another more concrete example of this is asking the question: 'What is the best X?', Any definite answer to such a prompt is a simplification and is not 'perfectly' representative of the underlying system.

[14] Usually when people refer to uncertainty, they mean perceived uncertainty of a system. I reframe uncertainty as inherent and the perception of it more separately. As an example: [Chollet, François. "On the measure of intelligence" arXiv preprint arXiv:1911.01547 \(2019\)](#)

[16] 'Useful' in an incredibly bare sense, a shorthand for "this just seems to work", or even worse: "it doesn't fail completely which I otherwise would've been aware of".

[17] Without making a distinction between the differences of 'ignorance' and 'naivety', for the purposes of my argument they are the same.

[18] [Hutter, Marcus. "Hutter Prize"](#)