



Autonomous Worlds (Part 1)

by [ludens](#)

A World is a container for entities and a coherent-enough internal ruleset about how they behave¹. When a system of entities and rules comes to life, it becomes a World.

We inhabit Worlds – both physical and conceptual. We learn how to work and behave within them. We engage in tribalism, spatial reasoning, and territorialism, even within Worlds that live entirely in our minds. We have a sense for the boundaries of Worlds and their rulesets.

Worlds exist within *books, games, social groups, and religions*. Amongst those, we can find *The World of Narnia, Christianity, and the Commonwealth Law*.

Worlds run on everything from letters to wikis, bedtime stories, constitutions, databases, and, most importantly, our collective human intelligence. To say that a World "runs" on *X* is to say that *X* is the reason for the persistence of the World, the reason why we continue experiencing the container as a World, as being *alive*.

Worlds sometimes live entirely within the minds of people, with some light physical footprints: books, computer memory, etc. However, the physical artifacts of such Worlds *are not* the reason that these Worlds are alive. Printing a million copies of a book doesn't create a World, unless people read it, care about it, and inhabit it.

Diegesis

To be precise when talking about Worlds, we need to define *diegesis*. Something is diegetic *if it is in the World*. And for something to be *in* the World, it needs to have respected the *introduction rule* of the World.

The notion of diegesis is important when defining the boundaries of Worlds. Remember, a World is a container.

Let's go through some examples to build up some intuition. We will use the word *entity* to describe any constituent of a World: event, character, rule, fact, etc.



The World of Harry Potter.

In the World of Harry Potter, the introduction rule is very simple: if an entity is included in a story written by JK Rowling and published under the Harry Potter series, it is diegetic. Otherwise, it isn't.

The World of the US Dollar

This World is alien to an average person, and so is its introduction rule. Its entities are things like *authorities*, *balances*, *debts*, and *values*.

The introduction rule goes as such: If an *authority* attests the existence of a balance or debt, it is diegetic. Additionally, if enough economically-powerful entities accept the "dollar value" of an entity – physical or not – its corresponding *value* becomes diegetic.

The World of Warcraft

In World of Warcraft, the introduction rule is formalised using computer code. If the game server relays the existence of an entity to players, it is diegetic. The truth value of statements about entities – like “my character is level 60” or “our guild is the best on the server” – is dictated by the C++ code written by Blizzard engineers.

Diegetic Boundary

Some Worlds do not have clearly defined boundaries, and certain entities can appear to be diegetic only to a subset of people.

Most Worlds don't suffer too much from ambiguous or nebulous diegetic boundaries. Others, like the USD World, are so important to our lives that we've decided to spend an immense amount of time precisely defining and enforcing their introduction rules and borders. You can think of bureaucracy and law as a form of gravity: they attract blobs of coherent entities together and define a strict boundary for what **is** and what **isn't** diegetic.

Formalising introduction rules using "law" and "code" has proved to be of utmost importance to the mission-critical Worlds permeating our lives. These tools have given Worlds *harder* diegetic boundaries.

Soft diegetic boundaries often involve authorities or social consensus as a form of introduction rule, whereas *hard* diegetic boundaries are enforced with clear transparent rules: law, code, or mathematics. Our top level physical World – the Universe – has a *very hard* diegetic boundary enforced through its introduction rule: Physics.



“Hard” or “Soft” boundaries can both be desirable features for a World. “Fan fiction” is the practice of playing with soft diegetic boundaries, while commerce requires a World with hard boundaries: doubts and arguments over the validity of someone’s tender hinders trade.

Blockchain: A technology for creating canon

Worlds have a tech tree: language, writing, law, and psychology are key discoveries that have enabled the creation of some of the most important Worlds in society.

In 2008, an email exchange introduced one of the biggest breakthroughs in World technology: Bitcoin.

Bitcoin is a Blockchain: a network technology used to produce consensus, or “canonicity.” In the case of Bitcoin, network participants reach agreement over the canonicity of a set of balances. Something being canonical is equivalent to an entity being diegetic from the lens of Worlds.

Bitcoin is a World. Just like the World of the US Dollar, it’s a weird one. The entities in the World of Bitcoin are balances and addresses, and the introduction rule is defined in computer code. Addresses have balances, and the introduction rule for spinning up new address-balance relationships goes something like this: part of an existing balance can be “spent” via a cryptographically-signed “transaction” to transfer it to another address. And – most importantly – the balance of an address can be increased through “mining,” an expensive computational process.

Blockchains are a type of substrate for Worlds. They unambiguously hold the set of *all* of their diegetic entities within their state. Additionally, they formally define an introduction rule with computer code. A World with a Blockchain substrate enables its inhabitants to participate in consensus. They run a network of computers reaching agreement on each introduction of a new diegetic entity.

There are two Blockchain concepts that are important to define from the perspective of Worlds:

1. A Blockchain *state root*

A **state root** is a compression of *all* entities in the World. With a state root, one can determine whether any entity is diegetic. Believing in the state root of a World is equivalent to believing in the World itself.

`0x411842e02a67ab1ab6d3722949263f06bca20c62e03a99812bcd15dce6daf26e` was the state root of Ethereum – a World with a Blockchain substrate – on the 21st of July 2022 at 07:30:10PM UTC. All entities of the World of Ethereum were taken into account in



the calculation of this state root. It represents the entirety of what was and what wasn't diegetic in that World at that specific time.

2. A Blockchain *state transition function*

Each Blockchain defines a **state transition function**. It can be thought of as an unambiguous introduction rule. It defines how we can change or introduce new diegetic entities, given both a set of inputs from people and machines, and the previous state of the World - the set of existing diegetic entities. In the case of Bitcoin, the state transition function defines how balances can be spent.

In a World with a Blockchain substrate, the *belief* of participants in the introduction rule entails **total acceptance** of the entities introduced by it. “Belief” here needs to be defined: An inhabitant of a World with a Blockchain substrate believes in the introduction rule when two statements hold:

1. They (or someone they trust) participates in or verifies the digital “consensus” of that corresponding Blockchain. Through participation, they can independently retrieve the **state root** of the Blockchain, which – as described above – is a compression of all diegetic entities in the World.
2. They believe the specific *consensus algorithm* of the Blockchain is operating properly. Blockchains are not magic: they create diegetic hardness, but they aren't free lunches. Various attacks and failure modes exist for each specific Blockchain implementation.

I'd like to insist that this is *not* a default property of Worlds with formalised introduction rules. For example, a flash crash at the Chicago Mercantile Exchange led to outcomes that were rejected by almost all traders as “invalid,” even if the introduction rule of the World – an order book matching engine – had been formalised with computer code. Belief in the proper operation of a Blockchain's consensus algorithm preempts various “what if” scenarios commonly found in other Worlds with formalised introduction rules, such as:

- What if someone changed the introduction rule without telling us, the inhabitants of the World?
- What if the introduction rule has been misinterpreted?
- What if some entities were introduced in a way that bypassed the introduction rule?

Through digital consensus, Blockchains create some of the hardest diegetic boundaries around.

Autonomy



Blockchains are of course not the only type of substrate for Worlds. Remember, Worlds run on everything from tribal songs to databases.

Yet, **Blockchains as a World-substrate bring a qualitative increase in the autonomy of their World.**

Each World ranks differently when it comes to autonomy: some Worlds have an introduction rule relying on the existence and participation of a permissioned individual to introduce new diegetic entities (eg: Harry Potter); others rely on the consensus of a group of people to interpret and enforce their introduction rule (eg: a national legal system, the World of the US Dollar); and some are in need of *untampered* computers running their formalised introduction rule (eg: The Chicago Mercantile Exchange, World of Warcraft).

In the limit case of a World's autonomy, no special individual or hardware is needed to introduce new entities and maintain the diegetic boundary.

Worlds with a Blockchain substrate are almost maximally autonomous: *anybody* can enforce the introduction rule, *without* damaging its objectivity. The disappearance or betrayal of any particular individual does not hurt the World: its diegetic boundary remains as hard as ever. Such Worlds can be nearly on par with systems like the English language, or physics itself.

Of course, autonomy is something you can only measure in retrospect. Before an actual existential threat faces the World, autonomy is often performative. Sometimes, a credible path towards autonomy is what allows Worlds to be seen as autonomous.

Autonomous Worlds

Given that "World with a Blockchain substrate" is quite a mouthful, we've started referring to these systems as **Autonomous Worlds**.

I like to think of Autonomous Worlds as similar to planets in our solar system, but *digital* instead of *physical*.

Think about Mars. Mars - with its mountains and ancient riverbeds, its complex geology, its thin atmosphere - is a World. Most of the time you cannot observe Mars by simply looking at the sky with the naked eye. Yet Mars is still out there, part of our solar system. If you were to use special instruments, you would be able to gather information about Mars, and this information would be the same for another person using similar instruments.

The telescopes used to observe Mars can be built by anyone. This makes it easier for us to agree on the fact that "yes, there is a big red sphere out there, and you didn't



make it up".

Additionally, the rocks and deserts on Mars keep existing if someone stops believing in their World. **Nobody can "unplug" Mars.**

Likewise, Autonomous Worlds have "telescopes" that anyone can build and use to reach consensus².

The entities of Autonomous Worlds remain diegetic as long as at least one person participates in the digital consensus. The introduction rule remains objective as well as transparent, and observing the state of the World is open to anyone with the right telescope. **Nobody can unplug Autonomous Worlds.**

Autonomous Worlds have **hard diegetic boundaries, formalised introduction rules,** and **no need for privileged individuals** to keep the World alive.

From Autonomous Worlds to Interobjective Realities

Thanks to Hilmar Petursson and Sina Habiban for inspiring this section.

In the book *Sapiens*, Yuval Noah Harari describes how, in addition to our shared *objective reality* (the Universe and its physics) and our *private subjective reality* (our own feeling and thoughts), we experience *intersubjective* realities: intangible concepts shared by multiple human beings. Prime examples of intersubjective realities are religions and money. Those realities – being subjective – have subtle different interpretations across people: *love*, an intersubjective reality, is experienced in very different ways. Even if shared, it remains intangible and subjective.

Autonomous Worlds enable "Interobjective realities". Through autonomy and an objective formalised introduction rule, we can reduce – or even remove – the (inter)subjectivity of those realities.

We have taken part in intersubjective realities for ten of thousands of years. Now, using the affordances of *autonomy* and *transparency* from Blockchain World-substrates, we can grant some of the rigidity and objectivity of our shared physical reality to our shared intangible realities. We can take the leap from intersubjective realities to interobjective realities.

Conclusion

In this essay, we introduced the concept of *Autonomous Worlds*: systems of entities and objective introduction rules that are enforced by blockchains. In a sense, Autonomous Worlds are *interobjective* realities.



In part 2, we will explore the role of Ethereum and smart contracts as substrates for Autonomous Worlds, and the idea of "Digital Physics": the large theoretical design space for rules that can be enforced by Blockchain or Smart Contracts.

Links and acknowledgements

Thanks to Sina Habibian, DC Posch, Josh Stark, Saffron Huang, Rafael Morado, Hilmar Petursson, Will Robinson, Lakshman Sankar, Arthur Röing Baer, Pillhead, gubsheep, gvn, and Nalin Bhardwaj for feedback on this essay.

Footnotes

1. We do not mean worldbuilding; which is focused on creating fantasy worlds in order to make fictional stories better and more consistent. [↩](#)
2. In Blockchain jargon they are called "full nodes". [↩](#)

