Truth, Trust, Facts, and the Blockchain

This essay was inspired a twitter conversation on the limits of blockchain in assessing the truth of a situation that may be timestamped on-chain.

CHARLIE MARKETPLACE

NOV 13, 2021

Share ···

Intro

This essay was inspired a twitter conversation on the limits of blockchain in assessing the *truth* of a situation that may be timestamped on-chain. Specifically, that placing information onto the blockchain does not in and of itself make that information “true”. The goal of this essay is to provide a framework for assessing whether a phenomena is “true” in a narrow sense. It begins with a strong assumption on the limits of *objective* “reality” and how a practical rational actor would combine [Facts] + [Trust] into a [Truth]- where truth is defined *subjectively* as what a practical, rational actor **should believe**.

On the Nature of Truth

Objective reality does not exist, because reality must be *perceived*. This is **not** to say anything about solipsism (“the view or theory that the self is all that can be known to exist.”). Here is a clarifying example.

Billy is blind. Although blind, Billy has several other perceptions- he can hear and interpret certain sounds as language, distinguish hot from cold, identify different materials by touch, among other perceptions. Alice explains to Billy that there is a perception she has, that he does not. She can perceive properties of objects without touching them, she calls this “sight”.

Having never experienced sight, Billy does not believe her. Our story could safely end here. What does Billy have to gain from believing her? Understanding sight won’t give him the perception. What does he have to gain from doubting her? Maybe for their friendship, he should just trust her.

Here is where I make the same argument in two different ways-

1. Billy should *selfishly* believe her, because having a friend with this perception may be beneficial to him in the future.

2. Billy should *practically* believe her, because *most people*, having sight, believe this perception exists and he will likely interact with many people holding this belief.

Separately, as someone with sight, I think Billy should believe her because **I know sight is real** and I think Billy **should believe real (**objective?**) things**. (Note- I don’t have to believe Alice has sight for me to believe sight is real- I trust my own perception of it. This is important). Generally, I think we all have this sense, that people should believe our truths.

Zero Knowledge Proofs

A Zero Knowledge Proof is sufficient evidence information is true, without revealing anything about the information beside that it is true. Alice and Billy can create a zero knowledge proof by agreeing on what evidence would be sufficient to convince Billy that sight exists (and optionally, that Alice has it).

Let’s say Alice and Billy get two balls. Alice tells him that the one in his left hand is “green” and that the one in his right hand is “red”. Billy doesn’t know what colors are and doesn’t care. He carves an indentation on the “green” one and a different indentation on the “red” one, so that he can tell them apart by touch. He trusts his ability to randomize which one is in which hand and moves away from Alice.

Alice says, “raise either ball and I will tell you which color it is without touching them. You will know the color based on the indents you made”.

Billy believes he can prevent Alice was touching the indentations he made and agrees.

They play 1 round. Alice correctly says “green” and Billy accepts he did raise the ball with the appropriate indent.

Should Billy believe sight exists? Should he believe Alice, has it?

Billy says, “You got lucky!”

Alice says, “I can do this with perfect accuracy as many times as you need to believe it”.

Billy says, “Ok, we’ll do it 20 times. That way it’d take 1 in a million odds for you to be right”.

After 20 rounds, Alice had 100% accuracy. Billy accepts that it is possible for someone to perceive properties of an object without touching them, he agrees to label this perception “sight”.

Because Billy trusts Alice and him were alone, he also believes Alice has this perception. But notice we haven’t proven that. It’s entirely possible Alice was blindfolded with an earpiece and someone else was feeding her the information.

This is the fundamental flaw of trying to prove anything as Truth through Fact alone. It can’t be done. There’s always some kind of possible exception or situation that could block us going directly from Fact to Truth. We have to *Trust* those exceptions or situations are not happening and that requires us to be both rational & practical. If Billy truly required the odds to be 0- that there was absolutely zero possibility Alice was getting lucky; Alice would have been there for an infinite amount of time.

Truth = Facts + Trust. The goal in assessing the truth, is to minimize trust and maximize facts. But this is strictly not possible- all we can do is set practical odds for us to be willing to believe some amount of facts (i.e., that sight exists) and then trust the possible exceptions didn’t happen (Alice did not have help) in order to state a truth (Alice has sight).

What the Blockchain CAN do

A blockchain transaction has the following immutable characteristics:

1. A settled transaction has an initiating address, i.e., it was requested by **an account**.
2. The transaction will have calldata, i.e., it contains **information**.
3. The transaction will have a block timestamp, i.e., it happened at a **specific time**.

This [Who] [What] [When] can serve as some of the facts around a phenomenon. It is a **necessary** but **not sufficient** set of facts useful in determining whether the information contained is *Truth*.

What the Blockchain CAN’T do

A blockchain transaction cannot tell us more than what it contains, that is:

1. It does not know which **person** was in control of the account at the time.

2. It does not know if its information is **factual**.

3. It does not know when its information was **created**.

These are important caveats and certainly should push us to believe blockchain has severe limits in identifying truth. But like Billy and Alice- if you combine a practical threshold to accept something as Fact; and a certain amount of rational Trust, you can form a belief believable enough to be Truth.

Social Consensus

The trust layer of a blockchain is its social consensus. If a sufficient number of people believe something to be true, such that the belief becomes useful, then given a practical threshold of fact and a rational amount of trust, we should believe that belief too in order to gain its usefulness. Here's three examples ranging from trivial to fundamentally important for Truth in the digital age.

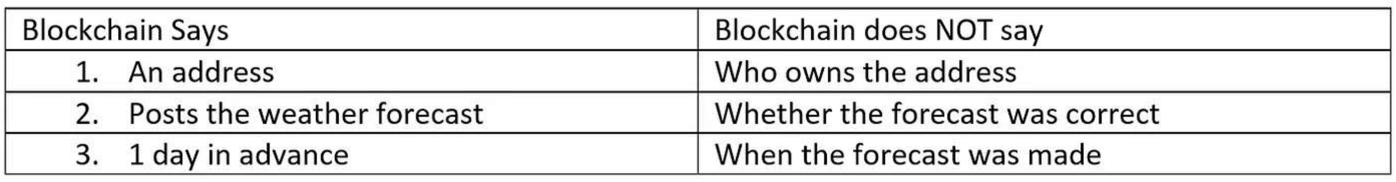
Example 1: Will it be hot tomorrow?

If it's hot tomorrow you'll pack shorts for your trip, otherwise you'll pack pants. How do you decide what to pack? Personally, I would just go to my *trusted* online weather forecasting company and believe them. But imagine you're an extremely skeptical person. What could you do to trust its tool?

You could check its previous forecasts against their reported actual temperatures. But then you'd have to trust they don't retroactively update their forecasts to look right.

The main solution would be to start tracking their forecasts versus temperatures you record, until you have a sufficient track record to believe them in the future (e.g., 20 days, just like when Billy made Alice guess the ball color 20 times). This is a lot of work though and you don't have time to start this experiment now.

The blockchain could help here. Let's imagine there's an Ethereum address that places a weather forecast for your target area directly on the blockchain 1 day in advance of the weather. These 3 facts combined with other minimally trusted non-blockchain information may allow you to believe the forecast and pack appropriately.



Blockchain Says	Blockchain does NOT say
1. An address	Who owns the address
2. Posts the weather forecast	Whether the forecast was correct
3. 1 day in advance	When the forecast was made

Let's accept that there's a 20+ day history of forecasts from this address. Then let's make a leap of faith that competing weather tools don't collude because they want to be known as the most accurate to take a larger market share (a very capitalist assumption that may not hold in other contexts). Finally let's say that there is a competing service that posts their same day temperature on the blockchain and we can use it as a reference. In order to trust this specific weather company, we'd want to do the following:

1) Confirm the address we're most interested in is controlled by the weather company.

2) Confirm the "same day" address is controlled by a non-colluding company.

3) Confirm the forecasts are made *at least* one day in advance.

4) Identify a reasonable threshold for trusting the forecast relative to the reference.

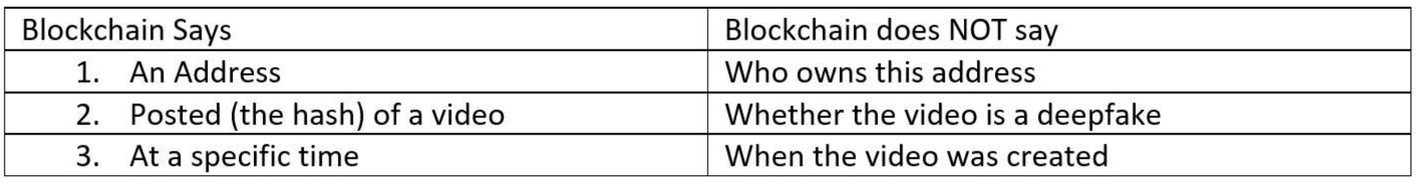
Zero Knowledge Proofs in the same structure as Billy and Alice can solve (1) and (2). We could ask the companies to repeatedly post our randomly generated choices on-chain near immediately to confirm (at least) they control the addresses that post the data. We could confirm (3) with the blockchain's factual timestamp. (4) We would need to define for ourselves what metrics to use and what kind of tolerance we have (maybe within 3 degrees at least 15 of the last 20 days).

We have the facts from the blockchain, combine it with trust (some earned through proofs, some contextual) and we get something we *should believe* because it's useful for us packing.

Example 2: Keeping a primary copy of a President's Speech

Deepfakes are already being used in politics to dangerous effect. There are numerous artificial intelligence companies out there attempting to identify if a picture or video was altered using the most well known deepfake methods. Blockchain can't tell us if a video is a deepfake, but what it can do is give us facts to lay our trust on for determining primary sources that *should be believed* to be true representations of the event.

Let's say President Biden makes a speech on January 3rd, 2022. It's recorded by CSPAN and broadcasted by 5 major mainstream media outlets along with numerous independent streamers (e.g., on YouTube). What can blockchain do to help us reduce the risk of believing a deepfake?



Blockchain Says	Blockchain does NOT say
1. An Address	Who owns this address
2. Posted (the hash) of a video	Whether the video is a deepfake
3. At a specific time	When the video was created

Let's accept there's a history of important political speeches posted by some addresses. Let's make a similar leap of faith as before, in that the competing media outlets seek to profit from different interpretations of events, but they have a minimally overlapping set of primary sources they use to make those interpretations (i.e., they are commenting on the same speech recording). Finally let's say there is a readily available means of confirming videos are the same or not (i.e., we can hash the data within a video file without a hash collision).

We'd want to do the following:

1) Confirm which address(es) are controlled by the broadcasters

2) Confirm the videos match the hash posted call data

3) Confirm the videos are posted *after* the speech is complete

4) Identify a reasonable threshold for video matching and time from broadcast to post that allows us to trust it is not a deepfake.

Zero Knowledge Proofs again can help us with (1); we can at least know which broadcasters have access to which accounts (we must trust they don't collude for external reasons). We can take the videos they make available (e.g., on their website) and confirm they hash to the value they posted using a hashing tool we trust and look at the transaction timestamp to confirm (2) and (3). For (4) we'd need to define metrics for ourselves but they may include CSPAN and other media outlets consistently outputting video hashes within 1-2 hours of a speech ending (limited time to deepfake), those video hashes matching each other, that subsections of the videos hash to the same value, and possibly that they align to independent recordings from YouTube.

With tools like Arweave and InterPlanetary File System (IPFS) we can store the video-hash pairs in a decentralized way and have a permanent record of the speech with facts from the blockchain and enough trust (some earned, some a leap) to consider it the canonical speech.

Example 3: Estimating COVID-19 deaths in the USA

How do you know someone died? There's an insane amount of trust needed to even know someone you haven't met lived. I remember where I was when I found out Michael Jackson died. I was an avid listener of his music. But how do I know he isn't with Elvis in Cuba?

Even starting a list of what you'd have to trust is overwhelming. You'd have to trust his birth & death certificates were actually issued by the issuing party at the time of issuance (we've seen blockchain can help with some of this with zero knowledge proofs of birth/death certificate issuers and times of issuance. Possibly along with hashes of key information like name, hospital, cause of death etc.). You'd need to trust the issuers of the certificates were correct in their issuance- that he was alive at birth and dead at death. You'd have to trust he wasn't swapped out with a lookalike at some point during his life- which some people do believe about celebrities- so we'd need some kind of consistent proof of identity (e.g., fingerprints or DNA tests that we'd have to trust).

So, what could blockchain do to help with a problem like tracking deaths in a country for a specific reason? And can it do it without releasing tons of private information? Especially knowing that up to 12,000+ people are currently declared dead incorrectly each year!

Imagine some amount of addresses are posting hashes of death data to the blockchain sufficiently close to the time of death. As previously stated, we'd need to know the relevant issuing party (e.g., the Social Security Administrations Death Master File team) controls the address; we'd need to be able to quickly confirm the inputs of known death data hash to the value posted onto the blockchain, and that the data is posted after the death but not so much farther that there is time to tamper with it (e.g., finding a hash collision to cover up Michael Jackson moving to Cuba). There are numerous possible structures besides using the SSA as a single source of truth.

One could imagine that competing (non-colluding) hospitals post (hashed) patient data to the blockchain, and that they are sufficiently consistent in not posting data about patients after a patient has received a death certificate from a facility (not only from itself). That causes of death are certified by a known examiner with a consistent history too; or that each death is tied to a COVID-19 test.

This problem is incredibly complicated and the blockchain's help here is significantly more limited due to the difficulties in verifying the information oneself (unlike checking a video hash, you wouldn't readily find lists of deceased people's names, birth dates, social security numbers, death dates, etc. for hashing).

But hopefully, reading through this, you could see how in a fully mature blockchain ecosystem, it could be possible to structure some baseline facts (certificate issuances, issuers, and time of issuance), combine them with a minimum amount of trust (competition, competency, avoiding hash collisions) to reach a truth to be believed (R.I.P. Michael).

Conclusion

Blockchain is not a panacea but being able to access credible timestamps and prove they are related to minimally trusted parties, allows the content of a blockchain transaction to inherit the trust of its party. Within credibly competitive contexts, the ability to quickly verify, compare, and contrast blockchain records provides a trust layer to the foundational facts, ultimately supporting people in figuring out Truth- the things they *should believe* because believing them is practical, rational, and ultimately useful.