On Bitcoin: A Study in Applied Metaphysics

This essay is dedicated to the memory of Katherine Hawley.[1]

Martin A. Lipman

Abstract: Bitcoin was invented to serve as a digital currency that demands no trust in financial institutions, such as commercial and central banks. This paper discusses metaphysical aspects of bitcoin, in particular the view that bitcoin is socially constructed, non-concrete and genuinely exists. If bitcoin is socially constructed, one may worry that this reintroduces trust in the communities responsible for the social construction. Although we may have to rely on certain communities, I argue that social construction doesn't imply a demand for trust because the relevant communities don't take on any relevant commitments. Bitcoin is indeed trust-free.

Keywords: bitcoin, cryptocurrency, social construction, social entities, applied metaphysics, trust, commitment.

Bitcoin was invented to be a digital currency that could be exchanged between parties without requiring trust in institutional intermediaries such as central and commercial banks. Although it started small, on the computers of a few programmers, currently countries around the globe deliberate on how to handle bitcoin. Some countries embrace it, others ban all forms of interaction with it.

This is an essay in applied metaphysics. I argue that bitcoin is a type of abstract substance (or stuff) that genuinely exists, and that comes in portions that we can quantify over and count. If bitcoin exists, it is naturally taken to be socially constructed in the sense that its existence and properties are due to social conventions. The second

[1] When Katherine supervised my PhD, we always spoke of issues in general metaphysics, such as persistence and change. The last time Katherine and I met, she had just come out of a meeting at *St Andrews' Centre for Exoplanet Science*, of which she was a member. This led us to talk about applied metaphysics as a way of enabling interdisciplinary engagements and societal impact. I had always been skeptical of the philosophical significance of applied metaphysics, but this conversation made me see things in a different light. So, when invited to contribute to this special issue, I decided to write an essay in applied philosophy. Writing this essay has made me think back to that conversation and Katherine's approach to philosophy: non-pretentious, insightful and always open to exploring new ground. I'm incredibly grateful for having known Katherine and for this opportunity to continue to learn from her. For some of Katherine's views on applied metaphysics, see Hawley (2017a). (I want to add, Katherine was no fan of footnotes (nor of writing in the past tense (nor of many side-comments in parentheses)) but, luckily, she did have a wonderful sense of humor).

part of this essay features a discussion of how bitcoin's being socially constructed is compatible with the original intention of serving as a trust-free medium of exchange.

The aim is to stay close to what might be a naïve or pre-theoretical view of bitcoin, to clarify it using some of the conceptual tools used in contemporary metaphysics and to make an initial case for the resulting view. The essay hopes to speak to two types of audience, namely, to those with an interest in bitcoin and its philosophical aspects, and to those with an interest in the general metaphysics of social entities. Maintaining accessibility for both audiences required that I include some basic explanation of bitcoin and of central metaphysical concepts.

There is no comparison with other cryptocurrencies. A general discussion of them requires more space than is available, given how much variation there is in their design, aims, history and involved communities. Similarly, there is no discussion of whether bitcoin is a form of money or not (on which see e.g. Passinsky 2020a), and little discussion of how the offered accounts bears on the many interesting normative questions raised by bitcoin (on which, see e.g. Bailey, Rettler, and Warmke 2021a; 2021b; forthcoming).

1. A short introduction to bitcoin

First a bit of history. Bitcoin's earliest mention was in comments on a mailing list by the pseudonymous 'Satoshi Nakamoto' in 2008. Satoshi Nakamoto published the first designs of bitcoin in a white paper, explained and improved it with the help of others, through email and on online message boards. Nakamoto went silent in 2010 and to this day, no one knows who Satoshi Nakamoto is.

Nakamoto thought that conventional currency is problematic because it requires people to trust a host of financial institutions. As he puts it in an oft-cited passage:

> The root problem with conventional currency is all the trust that's required to make it work. The central bank must be trusted not to debase the currency, but the history of fiat currencies is full of breaches of that trust. Banks must be trusted to hold our money and transfer it electronically, but they lend it out in waves of credit bubbles with barely a fraction in reserve. We have to trust them with our privacy, trust them not to let identity thieves drain our accounts. (Nakamoto 2008: 1).

Fiat currency is a currency whose issuance is authorized by governments and which isn't backed by any commodity, such as gold. For any conventional fiat currency, be it physical cash or digital forms of money, the monetary functions that we rely on requires a level of trust in financial institutions. For example, when we save, we rely on the saved currency maintaining value over time. How well a currency maintains its value depends, amongst other things, on how scarce it remains and hence requires a certain level of trust in the financial institutions that have control over the supply.

Increasing levels of trust are needed when it comes to digital forms of conventional currencies. Any digital form of currency relies heavily on a record, a ledger, of who has what amounts, since digital information can be easily copied and manipulated. When institutions hold the ledgers, any transaction (such as a payment) needs to go through these institutions and be approved by them. These financial institutions can block usage of a currency, or even seize funds.

Violations of the trust placed in financial institutions are not just hypothetical. Think of the real-life cases where, due to corrupt or irresponsible monetary policy, currencies collapse due to hyperinflation. Think of governments freezing accounts of protesters, political autocracies blocking funding of opposition parties or using digital currencies for financial surveillance. If one lives in the fortunate context where such malpractices don't arise, this may make it harder to see the ways in which human control over currency can be problematic (Gladstein 2022). But even in financially privileged contexts, our trust in financial institutions can become salient, such as when elevated levels of inflation erode purchasing power and financial stability.

Perceived violations of trust are clearly key drivers behind the creation of bitcoin and the early adoption by cypherpunk activists. It's no coincidence that bitcoin emerged in 2008, at the time of the Great Recession and the bailout of banks around the globe. The first block in bitcoin's ledger – the "genesis block" – contains a reference to a newspaper article on the bailout of banks, leaving little doubt about the reasons behind its creation.

So, bitcoin is to offer a digital value-bearer that is stripped from the influence of the financial institutions that conventional currencies rely on, such as central and commercial banks. The design aims to do without the typical knobs and levers that can come under central control by some group or institution, using instead decentralized networks and cryptographically secured communication amongst various components in the system.

How does this work? Bitcoin relies on a publicly available ledger, that records transactions between so-called addresses. An address is associated with a string of letters and numbers. Each address comes with a private key, a kind of password. Anyone who knows the private key of the address can spend any bitcoin associated with (or 'on') the address, and hence is naturally said to own it. If the ledger's latest state records a transaction of 0.1 bitcoin to your address and, say, no transaction from your address to another, this means that you can spend up to 0.1 bitcoin from that address to another.

Instead of the ledger being held by some designated entity, such as a bank, copies of bitcoin's ledger are held at so-called *nodes*, computers that run software that constantly downloads the newly updated ledger and uploads it to other nodes in the network. Besides maintaining a constantly updated version of the ledger, nodes check new incoming transactions against a specified set of rules and transmits the valid transactions to other nodes in the network. Importantly: that a transaction is checked

and transmitted to other nodes does not yet mean that it's written into the ledger, it's initially just distributed across the network.

Writing valid transactions into the distributed ledger is the job of so-called *miners*. A miner is nowadays typically a computer that is optimized for mining only, but ordinary desktop computers were used in the early days. The miners compete to solve a math problem. Because these math problems can be solved only by trying out arbitrary solutions to it, any miner has some chance of solving it. Of course, the quicker a miner can try out solutions (the more processing power it has), the higher its chance of finding it and being granted the chance to write the new valid transactions into the ledger.

When a miner wins, the rules allow the miner to write a special transaction into the ledger (a 'coinbase transaction'). The rules allow the miner to add a fixed amount of bitcoin to their own address, hence increasing the total supply of available bitcoin, together with the valid transactions between other addresses, broadcasted by the nodes. In this way, the miner is rewarded with some bitcoin for its supplied processing power. This happens roughly every 10 minutes. The transactions written in the ledger come in added blocks, and the ledger takes the form of a chain of blocks, a 'blockchain' with a history of blocks of recorded transactions. The updated ledger is transmitted again to the network of nodes, which use it to check newly submitted transactions again for validity. And so it goes.

The total supply of bitcoin only increases through these rewards to miners. Every four years the fixed rewards halve. As things stand, this process ends in 2140, when 21 million bitcoins will have been issued. After this the supply of bitcoin no longer increases. Whenever you send around some bitcoin, you pay a fee, which also go to the miners, and further incentivizes their contribution of computational work.

This brief introduction is simplified and leaves out a range of complexities, but it should suffice for our purpose (see Warmke 2021 for further introduction; and Antonopoulos 2017 for a detailed explanation).

2. *Ontological realism: on bitcoin's claim to existence*

The natural first question for a metaphysician is whether there is any such thing as bitcoin. Do I really give you *something* when I send bitcoin to your address? Should we think of this bitcoin that 'changes hands' from me to you as genuinely existing?

In asking this question, we first need to make a distinction between the bitcoin network on the one hand, and the bitcoin that we say are owned by people and exchanged between people. We are only concerned with the latter, the bitcoin apparently owned and transferred.

A basic argument for the existence of bitcoin appeals to the fact that we can own, receive, and give bitcoin. Say you own bitcoin. You can only own things that exist, that

is to say, you cannot own what doesn't exist. So, there exists something that is the bitcoin that you own. Similarly, you can only give away things that exist (that is to say, you cannot give away what doesn't exist). Since you can give me a bitcoin, there exists something that is the very bitcoin that you can give me. (Compare the more general argument in Passinsky 2020b: 432).

We can call this an argument from existence-entailing properties and relations. The instance I'm putting forward here has the following shape: (premise 1) we can own bitcoin and transfer bitcoin to someone, and (premise 2) if we can own or transfer bitcoin, then bitcoin exists. I'm not under the illusion that this settles the matter. The argument is a starting point, not an endpoint: it helps us proceed a bit more systematically in our evaluation of the claim that bitcoin exits. The conclusion that bitcoin exists can be resisted if either one of the two premises is resisted, so if we either do not truly own bitcoin, or if our owning bitcoin doesn't imply its existence.

Start with the second premise: we can ask ourselves how plausible it is in general that, if someone gives *x* to someone else or owns *x*, then *x* exists. Let me focus on the question of owning something (as the relevant sense of 'giving' is arguably tied to the idea of a change in ownership). We would not normally take ourselves to own anything that we believe not to be there. One cannot own a pet unicorn, for instance. You can pretend to do so, or imagine owning one, but you cannot truly own one. If I tell you in all seriousness that I own a unicorn, you will take me for mad, presumably because it is generally understood that this would imply the existence of this unicorn and generally understood that there is no unicorn.

Someone might try to object as follows: fictional characters can be protected by copyright laws. If, for instance, DC Comics has the exclusive right to make movies about Batman, should we not say that DC Comics owns Batman and hence owns something that doesn't exist (namely Batman)?

The copyright law gives exclusive right to create creative works, such as comic books and movies, about something. It's less clear whether we should think of this as implying ownership over a fictional entity. But even if it did, I don't think this casts doubt on the argument. When we consider fictional characters, we should distinguish the entity that is created by an author at a particular point in time, a fictional character that is used in all sorts of ways in movies and comic books, and the *would-be person* Bruce Wayne that lives in Gotham and drives around at night in a black suit (Kripke 2011). The author doesn't create a person that drives around at night in a black suit. There is no such person. What is brought into existence by an author is Batman-the-fictional-character (which is plausibly an abstract object), what is created is not Batman-the-person-who-lives-in-the-city-Gotham. It's the fictional character, which indeed exists, that would be owned by DC Comics, if anything is. There is no counterexample here. Ownership implies existence, even in cases like these.

Consider the other premise: how good is the claim that we indeed truly own and give bitcoin? The reason for focussing on the ownership of bitcoin is that facts about

ownership are social facts: there are social patterns that suffice for facts about ownership. This means that we can look to social behaviour and conventions when evaluating if somebody owns something, we have an independent handle on this question. When we take this approach regarding bitcoin, the claim that people truly own bitcoin seems in a fairly good epistemic standing, at the very least compared to other matters that we normally assume we own. Bitcoin is embedded in the social patterns in the way one expects if there is genuine ownership. Across the globe, there are enough people willing to exchange goods and services for bitcoin to think that the claim that one owns bitcoin is as good as any similar claim about owning a piece of land, a house, a certain amount of money, or some other asset. The more people are disposed to treat the ownership of bitcoin in the way that they treat the ownership of ordinary money or other assets, the less credible it becomes to speak of 'true ownership' in the ordinary cases and 'pretended or fictional ownership' in the case of bitcoin. Bitcoin is taken seriously across a wide variety of communities around the world, from professional and amateur investors to 'unbanked' communities and those suffering from hyperinflating currencies. One's bitcoin falls under tax regulation in many countries. For these laws, one owns bitcoin just as much as one owns a house, a boat, or a stock. Finally, those who own bitcoin do not generally see their owning bitcoin as an act of pretence and would emphatically deny that it's a mere game.

In short: the social patterns seem to suffice for there to be facts about who owns which bitcoin, and so the claim that we genuinely own bitcoin seems as reasonable as claims of ownership that are widely accepted, such as that of owning a house or the money on one's bank account. Those who deny that we truly own bitcoin (perhaps precisely because they insist that bitcoin doesn't exist) would be committed to an ad-hoc gap between the relevant social patterns on the one hand, and the social facts of ownership for which they normally suffice in other cases.

Further clarifications can help us see the tenability of ontological realism about bitcoin and dispel some initial worries. First, claiming that something exists doesn't imply that it's somewhere to be found in space and time, or that it's a material entity. Whether something is to be found in space and time if it exists depends on what kind of thing it is, and isn't built into the very notion of existence itself (Quine 1948: 23). Bitcoin isn't the sort of thing that must be found in space if it exists. Bitcoin has a good claim to being an abstract object. Now the notion of being 'abstract' is known to be unclear (Lewis 1986: sec. 1.7). I use 'being abstract' as shorthand for being necessarily non-material and non-spatial. Bitcoin is not plausibly material (what would it be made of?) and it is not plausibly spatial (where would you have to point, to point to the location of your bitcoin?).

When we say that something abstracts exists, we are just saying that it's there and that it stands in actual relations and bears actual properties. As I mentioned, ownership and related notions are matters on which we have some independent grasp, given that certain social patterns can suffice for the obtaining of ownership. But there are other properties that we are inclined to attribute straightforwardly to bitcoin. For instance,

given the current state of the blockchain, we can see that there must be a total of 19113 million bitcoin that has been put into supply so far. The property of *being such that a certain amount of it has been put in supply* is plausibly instantiated by bitcoin, and again: being an instantiator of any property of this kind suffices for it to exist, given our minimalist understanding of what it is to exist.

Secondly, claiming that something exists doesn't imply that it exists independently of human beings, nor that we only recently found a way to refer to these abstract things that were already existing (which one might say about other abstract objects, such as numbers). The more reasonable view is that bitcoin is socially constructed and started existing only when the requisite social conventions emerged. Many entities exist due to human beings, in some sense or other. Think of money, governments, nations, courts, universities, married couples. Socially constructed entities come into existence and are sustained in existence due to various contingent social phenomena, which may involve some type of collective action, social conventions regarding it, socially entrenched explicit attitudes towards it, or wider socially entrenched dispositions that involve it. I will use the term 'social patterns' for these.

That bitcoin is socially constructed and hence depends on such social patterns doesn't imply that it doesn't truly exist (compare Thomasson 2003: sec. 3; Mason 2016: sec. 4). On the contrary, being socially constructed is another candidate for an existence-entailing relation: for anything to be socially constructed is for it to have come into existence somehow based on, or due to, social factors, such as practices or intentions, and hence implies the existence of what is socially constructed.

Thirdly, it may be clarifying to briefly contrast the proposed ontological realism with a closely related alternative view that has been proposed in the literature. Warmke (2021) offers a fictionalist account of bitcoin, according to which bitcoin's ledger, the blockchain, is a kind of 'digital book', a kind of fiction, co-authored by all and everyone who transacts bitcoin. The ledger represents the fictional movements of the fictional bitcoin across addresses, and is a merely intentional object, presumably in the sense that it's merely represented as being there having various properties (Warmke 2021: 36, 37).

This fictionalism about bitcoin contrasts in various ways with the ontological realism sketched here. Warmke argues that the mere claim that bitcoin is fictional doesn't of itself imply that bitcoin doesn't exist, and that whether this is so depends on further assumptions about whether fictional entities exist or not (Warmke 2021: 23). Indeed, the main contrast concerns the way we think of the properties instantiated by bitcoin. Although both fictionalism and social constructivism see a dependence on human factors, the key difference is that a social constructivism about some *a*'s being F implies that *a* really instantiates the property of being F, whereas a fictionalist about *a's* being F endorses merely that *according to some fiction (a story, game or coordinated acts of pretense)*, *a* is F. The difference is subtle but matters.

According to fictionalism, most of the apparent properties and relations ascribed to bitcoin are merely properties and relations that bitcoin is *represented as having* by the relevant community. According to ontological realism, bitcoin genuinely has the properties assigned to it and hence the relation between the blockchain and bitcoin is not appropriately seen as one of representation (by the blockchain) and the represented (bitcoin). To represent that *p* doesn't make it be the case that *p*, whereas, according to ontological realism, bitcoin's blockchain being in appropriate states makes things be the case about bitcoin (as also emphasized by Glazier 2021, and further discussed below). For instance, the current state of the blockchain makes it be the case that there is currently a total of 19113 million bitcoin.

In the case of bitcoin, I assumed that facts of ownership are social facts, meaning that required social patterns suffice for the obtaining of facts about who owns what, and I argued that the required social patterns are in place that suffice for genuine ownership of bitcoin, and that this implies its existence. I also argued that there are no independent indications that people are engaging in pretense (they would typically disavow that it's merely fiction or pretense), nor are there indications of people generally treating bitcoin as merely intentional objects, like I would treat my imagined pet unicorn.

Of course, one may come to this discussion with a prior view that there only exist material objects and that there are no abstract objects *whatsoever*, or with the broader view that all talk of ownership is misguided as there isn't truly such a thing as owning anything, or that there is no such thing as social construction and that social entities are always only things that we merely pretend to exist. Nothing in this section can answer such wider inhospitable views, or convince anyone to abandon them.

The 'arguments from existence-entailing relations' are best understood as conditional on a view that admits socially constructed and non-concrete objects, and which has an appropriately minimalist understanding of what it is to exist. Given this limited scope, the argument may seem simplistic or 'thin', but such arguments can be useful in determining reasonable default positions, establishing that it's reasonable for someone to assume that bitcoin genuinely exists in future theorizing on the topic.

3. Socially constructed to be as it is conceived to be: abstract stuff indeed

When we're engaged with the metaphysics of something, we are not only interested in the existence of it, we are also interested in how something is individuated and what sort of ontological categories it falls under. I already touched on this when I suggested that bitcoin is a type of abstract object, but there is more to be said.

There are different approaches that one could take to this question. There is a widely accepted methodological dictum in metaphysics that we shouldn't read our ontology off our language use, a general methodological view that I very much share. Yet I think applied metaphysics cannot always blindly follow the methodological approaches we

take in more general discussion within metaphysics (compare Hawley 2017a: 177). In the case of bitcoin, we encounter a case where the apparent metaphysics underpins a methodological approach that goes against the general dictum. If bitcoin is indeed socially constructed, what is conventionally or generally assumed to be the properties of bitcoin may on that very basis be made to be properties of bitcoin. If how we talk and think of bitcoin shapes to some extent what it is that we postulate and attribute to bitcoin, then it makes more sense to treat how we talk and think about bitcoin as a (fallible yet informative) guide to the relevant ontology in this particular case. When the scientific understanding of the bitcoin's potential roles and effects within society develops, this will naturally serve as providing complementary guidance (Hawley 2018), again given that this embedding in society is part of what shapes bitcoin. An added advantage of treating how we talk as a (fallible) guide to the ontology is that our theorizing remains in touch with the language of public discussion.

As Warmke also notes (2021: sec. 6.1.1), we use the term 'bitcoin' in a variety of ways. The term 'bitcoin' is often used as a mass noun ('she has *some* bitcoin'). There is a closely related use of 'bitcoin' as the unit for *how much* bitcoin is sent or owned ('she has 0.2 bitcoin'). This all suggests that we think of bitcoin as if it were a kind of non-concrete 'stuff' or 'substance' that can come in different quantities. There is a fact of the matter about *how much* there is at a time and it makes sense to speak of bitcoin as being divided into portions (e.g. I can give you half of my bitcoin). Besides the mass noun use of 'bitcoin', we also occasionally use 'bitcoin' as a count noun ('she has *two* bitcoins'). This is naturally taken to refer to the 'portions of bitcoin stuff' which we can count and quantify over.

It's convenient to theorize about stuff indirectly by theorizing about the behaviour of portions of stuff (see e.g. Markosian 2015), so let us focus on the portions. When we distinguish the bitcoin that Alice owns from the bitcoin that Bob owns, the distinction is implicitly between portions of bitcoin, and hence it's a distinction between two abstract objects. These portions are distinct, amongst other things because of the different relations they stand in: the one is owned by Alice, the other by Bob.

There are good reasons to think that portions of bitcoin cannot survive a change in address. This underwrites bitcoin's economically important feature of fungibility. When something is perfectly fungible, equal quantities of it are always interchangeable, guaranteed to be of equal value. Two gold bars of 1kg are in principle interchangeable; two 1$ bills are interchangeable; and so are two barrels of oil. Compare this with diamonds: diamonds of equal quantity may not be interchangeable when they differ in how they are cut. When gold bars get identifying numbers, this can harm their fungibility as they now have a unique history and a way of tracking identifying properties that can start to bear on their valuation; for example, when you have the choice between a bar that has been mined in an environmentally responsible way and a bar that is known to have been confiscated illegally in the past, someone may value the first bar more than the second (so that one is willing to pay a premium for it) and could resist interchanging one for the other.

Bitcoin is fungible. Warmke (2022) explains very well why this is so. I will give a simplified explanation. Blocks in the ledger record transactions to addresses. Say that there are two blocks with transactions to address A, one transaction of 1.0 bitcoin from address X to A and one of 1.0 bitcoin from address Y to A, and no further transactions anywhere in the blocks of the ledger to A, nor any transactions away from A in any of the current blocks. This is what constitutes there being 2.0 bitcoin 'on' address A. As explained above, the record of the two prior transactions to A is what allows for the addition of a future block with a transaction of up to 2.0 bitcoin away from A to some other address. Say the owner of A creates a new transaction of 1.0 bitcoin away from address A to address B. The transaction will be recorded as simply '1.0 bitcoin from A to B', and this will be accepted by the nodes, given the prior transactions to A, from X and Y. Crucially, what is recorded are only the transactions of certain amounts, there is nothing that identifies the bitcoin that is sent to B as the very bitcoin received from X, or from Y, or as consisting of a mix of the bitcoin received from both, and this is likely intentional (as noted in Warmke 2022).

There is no way to track the portion of bitcoin that you have on your address through a history of transactions back to its origination, which one would assume to be the reward to a miner. Recorded transactions only specify *how much* goes from one address to another, not which portion of bitcoin.

In coming to a metaphysics of this, one could in principle postulate surplus ontological structure, going beyond what is fixed within the blockchain. This could result in facts about the individuation of portions of bitcoin across addresses involved in transactions, but doing so seems objectionable insofar as it goes against the features of the design that ensures the fungibility of bitcoin. The lack of identifying information tracking bitcoin across transaction histories is a feature that is desirable for what bitcoin aims to be, given that fungibility is economically important for a medium of exchange and these design choices make bitcoin fungible, and are likely intended. We should not attribute structure that something was intentionally designed to lack, we should treat the engineering and design choices as evidence for the ontological structure of a socially constructed entity.

Considering this, the following seems the most fitting ontology. When portions of bitcoin are sent to an address, the result isn't a compound within which these portions of bitcoin can still be identified. Portions of bitcoin are individuated by addresses. There are exactly as many portions of bitcoin as there are distinct addresses with a non-zero amount of bitcoin on them. A portion does not remain the same across a change in address. Different address, different portion of bitcoin.

Besides taking portions to be individuated by their address, we can further individuate them by how much they are, a standard assumption about portions of stuff. This means that adding two bitcoins to a single address creates a new single portion of bitcoin. When two distinct portions of bitcoin are sent to a new address, they stop existing, and a new portion of bitcoin emerges, associated with the receiving address. A portion of

bitcoin cannot survive a change in quantity, it cannot change in how much it is. Different size, distinct portion of bitcoin.

The proposed picture implies that a transaction doesn't consist of one and the same portion of bitcoin 'moving' from one address to another. A transaction is better understood as a transaction, not of the portions themselves, but of a certain *quantity* of the abstract bitcoin stuff, of which they are portions. We say that *some* bitcoin, a certain quantity of bitcoin, is sent from one address to another. The abstract stuff moves from one address to another, and in doing so, portions of bitcoin go out of existence as new portions come into existence.

A little mental model may help clarify the resulting ontological picture. Think of the addresses as labelling points on a flat grid, with bubbles sticking out on the addresses with non-zero amounts. The size of the bubbles of fluid on the grid are in proportion to how much bitcoin there is on the address. There is only ever one bubble at any point on the grid. One can open a channel between any two points on the grid, allowing the fluid to flow through. When you send some of your bitcoin away, we can think of this as opening a channel to another point on the grid and the fluid moving through the channel to another, where a new bubble emerges. As one bubble is destroyed, another bubble emerges somewhere else. The bubbles are analogous to portions of bitcoin, the fluid to the abstract bitcoin substance. The patterns in the emergence and destruction of bubbles on the grid correlate with movements of bitcoin-stuff between points on the grid.

Note that the current picture sees two ontological facts involved in a transaction: facts about movements of quantities of bitcoin stuff are necessarily correlated with patterns in the distribution of portions of bitcoin across addresses and yet, for some bitcoin to move from one address to another *is not just* for some portion to get destroyed and another portion to emerge, it's rather for some abstract stuff of a certain quantity to change address.

The sketched ontology allows us to maintain the existing ways of talking within our theorizing: when 'bitcoin' is used as mass noun, it refers to bitcoin-the-abstract substance, when it is used as a count noun, it refers to portions of the abstract bitcoin stuff (typically of the size of 1.0 bitcoin). From here on, I use 'bitcoin' to refer to the abstract stuff and this will be the focus in the remaining discussion. I explicitly use 'portions of bitcoins' to refer to portions of bitcoin, to avoid confusion.

4. More on the social construction of bitcoin

I argued earlier that the relation between the blockchain and features of bitcoin isn't one of representation. This raises the question of how we should we think of this relation. I want to propose that bitcoin's ledger is the basis for the conventions that govern the social construction of bitcoin. Bitcoin arises from blockchain-based conventions.

Social entities, and facts about them, depend somehow on communities. An influential account of social construction, first proposed by Searle (1995), takes collective intentionality to underwrite constitutive rules – conventions – that impose functions on pre-existing things. An example is a paper dollar bill: we collectively accept the rule that a paper bill issued by the Bureau of Engraving and Printing is money. This is understood to be the basis for a social postulation: by our collective acceptance of this convention, such bills *are* money.

Searle's account is restricted to imposing functions (or a status) on pre-existing material things, a restriction that has been rightly criticized (e.g. by Smith in Smith and Searle 2003). Building on the work of Searle, Thomasson (2003; 2009) expands the framework to make room for rules that allow for the collective postulation of new social objects, instead of merely making existing things fall under new kinds.

A common way of stating the relevant rules is in terms of collective acceptance (compare e.g. Thomasson 2003, 282), so that they take the following form:

> We collectively accept that if conditions C obtain, then there is some *x* such that it has feature F.

On this formulation, we collectively accept that, under certain specified conditions, there exists an entity with certain kinds of features.

As it is stated, this is a description of what people collectively accept. If we are to allow for the possibility of collective mistakes in what we accept, mere acceptance shouldn't as such necessarily suffice for the existence of things that are that way. For example, we could in principle collectively accept something that would attribute incoherent properties to a social entity. There must be a distinction between collectively accepted matters that makes things be a certain way and those that fail to do this. The mere identification of what is collectively taken to be a certain way isn't enough to ensure that there exists the relevant object with the relevant features.

One response is to assume that the social patterns that give rise to successful postulation should not be seen in descriptive terms, for example, perhaps the active postulations can be captured as a set of imperatives or instructions, of the form "if A, let there be an F!" (Fine 2005; 2012). I prefer to stay with a descriptive approach on which the ontologically relevant conventions are not stated in terms of something that is collectively accepted, but more directly as descriptions of acts of successful postulation:

> We collectively postulate that if conditions C obtain, then there is some *x* such that it has feature F.

We understand postulation to be a success case, so that if we collectively postulate that if A, then there exists an F, then it follows that there indeed exists an F if A. Of course, *we* as theorists aiming to identify acts of social construction can be mistaken. For instance, what might seem an act of successful postulation can turn out to be

merely a widely shared belief that is false. Further theorizing is needed to account for the difference between collectively accepted matters that successfully postulate, and those that don't, but this is not something I can get into here. The aim here is not to offer a reductive account of social construction, but to discuss its application to bitcoin.

Applied to the case of bitcoin, the relevant conditions involved in the postulational conventions are states of the blockchain. One key convention is the following:

> *Bitcoin existential rule*: we collectively postulate that if the bitcoin network is live, and bitcoin's blockchain includes transactions, then there is some *x* such that *x* is bitcoin.

Given that bitcoin network is indeed live, and new transactions continue to be recorded in bitcoin's blockchain, bitcoin exists.

If this much is right, we can expect closely related postulational conventions for various features of bitcoin (which we could call blockchain-based features) and whose instantiation consists in socially constructed facts about bitcoin. One example is how much bitcoin there is:

> *Bitcoin supply rule*: we collectively postulate that if the bitcoin network is live, and the current state of bitcoin's blockchain includes a total of unspent transaction output of *n*, then there is currently *n* bitcoin.

Similarly for facts about successful transactions, about what constitutes ownership of bitcoin (and portions of bitcoin) by people, and so on. We should expect there to be postulational rules governing various things, based in states of the blockchain.

On this picture, Nakamoto did not create bitcoin simply by writing the appropriate software and letting it run; he only created the bitcoin software this way, not the bitcoin we own and transfer. There was the (presumably implicit) act of establishing postulational conventions, which would be made explicit as: "let there be bitcoin and let it be governed by such and such rules based on states of the distributed blockchain". Such conventions later turned out to be postulation given the social entrenchment that followed, which started on the online fora and through email. We can imagine Nakamoto writing in messages to Hal Finney, one of the first known cryptographers to interact with bitcoin: "I have now sent you some bitcoin". Hal Finney endorses the convention and accepts the code *as* underwriting the transaction of some bitcoin. From this and further spreading of behaviour and mental attitudes, the social patterns came to postulate bitcoin and various (social) facts about it. Nakamoto did not just engineer code, he sparked the social patterns from which there came to be bitcoin.

5. *Socially constructed, yet trust-free*

The claim that bitcoin is (and continues to be) determined through social conventions raises a question about whether this reintroduces the sort of trust in human groups that bitcoin is meant to avoid.

As we saw, Nakamoto's aim was to create an 'electronic payment system based on cryptographic proof instead of trust, allowing any two willing parties to transact directly with each other without the need for a trusted third party' (Nakamoto 2008: 1). To this day, the central innovation of bitcoin is widely thought to have something to do with its minimized demands on trust. Contrary to standard financial systems, bitcoin is often assumed to be a trust-free financial system that requires users to rely only on code, hardware and the direct communication amongst parties enabled by the internet. The suggested image is that of a system that lacks the sort of human influence that normally calls for trust. This rough image of bitcoin may seem at odds with the metaphysics sketched above, which sees constant human influence and ongoing social construction, determining and sustaining many of the central features of bitcoin.

Trust is talked about in many ways, and we are not always very precise in how we use the term, not just in everyday settings but also across academic disciplines. Jacobs (2020) discusses how different accounts of the nature of trust create not just different conceptions of the role of trust in bitcoin, but also much crosstalk and confusion. As he sensibly suggests, we can only make progress by explicitly laying down how we understand trust and proceed our discussion from there.

Within philosophy, there is a widely endorsed distinction between trusting someone and merely relying on something or someone. The distinction is reflected in the kind of reactive attitudes we have to trust and reliance (Baier 1986). When you trust someone to do X, but that person doesn't do it, this reflects badly on the trusted person, and you feel wronged. You may demand an apology, or even feel betrayed. In contrast, when you rely on someone or something to do X and this person or thing doesn't do it, this just means that it would have been better for you if you hadn't have relied on this person or thing to do X on this occasion. We can rely on things, not just people. I rely on my computer to work. When it doesn't work, I don't feel wronged by it. The focus of our discussion is in the first instance on trust, not mere reliance.

I leave it open whether the trust in a group or institution, like the central bank, is not better understood as trust in the individuals that make up the group and execute the tasks of the institution. I do assume that the distinction between reliance and trust applies just as well to groups (although, as just mentioned, this might ultimately boil down to the distinction applied to the members that make up the groups; for discussion see Hawley 2017b: 247).

Besides the distinction between reliance and trust, I endorse Hawley's specific account of trust. Hawley (2014; 2019) proposes that to trust someone is to rely on that person to fulfil a commitment. Inanimate objects do not make commitments; they just do or don't, and hence cannot be trusted, only relied upon.

This account rightly predicts that ordinary fiat currency comes with a demand for trust in some groups. When storing and transacting using a standard currency, you are aptly described as relying on the banks (or, perhaps more precisely, the relevant bankers) to fulfil a range of commitments, such as a commitment to letting you transact freely in the future and a commitment to ensure that the currency maintains its value. When the relevant institutions don't fulfil this commitment, we feel wronged or betrayed, precisely because the banks are taken to be committed to doing these things.

One final preliminary clarification: we need to distinguish carefully between having to trust some group to do something with the help of bitcoin *when interacting with the group* and having to trust some group *merely based on interacting with bitcoin*. There exist many sub-communities that aim to establish principles for how one ought to interact with bitcoin and which push for all sorts of (political) aims and visions of how bitcoin may change the world; indeed, this started early with the enthusiastic reception of bitcoin within the cryptoanarchist and cypherpunk communities. It seems evident that sub-communities can take on some commitment, for example to push for a certain societal change *with the help of bitcoin*. That is a commitment to *use* bitcoin for some or other purpose and of course one can rely on such a sub-community to fulfil their commitments, and hence trust them. Much the same applies to the various companies that emerged around bitcoin, such as the many exchanges and apps. But these are independent acts of trust that are irrelevant to our discussion. One doesn't come to rely on such a group just in interacting with bitcoin, that is, when holding, sending, receiving, or mining bitcoin. Our discussion is about whether bitcoin itself (as opposed to the organizations and sub-communities revolving around it) demands trust in certain groups.

There are two natural candidates for communities of which one could think that they need to be trusted when interacting with bitcoin, namely the community of miners, and the broader and more loosely defined 'bitcoin community'. Let us consider these in turn.

We earlier used 'miner' for the machines dedicated to search for blocks but let us now use 'miner' for those who own and control these machines ('mining rigs') and decide how to use them. Does someone take on a commitment when she turns on a mining rig? I do not see what this commitment would be. The community of miners dedicates processing power that they control and let their machines search for a block, hoping to get lucky and earn bitcoin. They make no promises and do not undertake any tasks, it's a purely self-interested and opportunistic affair, much like panning for some gold along a river with some equipment.

Sometimes miners need to make decisions about changes to the code of the bitcoin software, and one might wonder whether this implies some kind of commitment on the part of the mining community to make the right decisions. The relevant bitcoin software is entirely open source: anyone can see exactly how it works and some can suggest changes or improvements to this code. Such suggestions for changes take the

form of so-called Bitcoin Improvement Proposals ('BIPs'), which are changes to the code that require endorsement by the miners to be implemented.

When some miners do not accept the changes to the rules, the blockchain branches, with each branch endorsed and continued by a different group of miners. These are so-called hard forks. (This happened during a dispute about the size of the blocks in which the transactions are written, the 'Blocksize War', see Bier 2021). When faced with these real-life fission cases, the question arises, which branch of the blockchain is bitcoin's? Or, using the account of social construction above, which branch becomes featured in the conventions responsible for the features of *bitcoin*?

Social patterns take a leading role here, not the miners. Conventions emerge from a complex mix of community discussion, decisions made in the surrounding infrastructure of exchanges and apps, as well as the market. The involved community includes those who own mining equipment, but also many who don't. It's the independent social conventions that determine which branch is (or will be) bitcoin's, what features bitcoin has and whether a given miner continues to contribute computing power to bitcoin or to something else. When there is a bitcoin improvement proposal, miners can use their processing power for a branch that incorporates the proposal or for a branch that doesn't – but they cannot wrong anyone in doing whatever they do.

This order of things—social conventions before miners—is relevant to bitcoin's resilience to attacks on the blockchain. Brute computing power can't force the hand of social conventions. Social conventions could in principle decide that a branch created by an attacking army of miners is not bitcoin's blockchain, even if it were to have more processing power behind it. No army of miners can force the direction of social patterns.

Let us consider the other suggestion, namely that trust is redirected to what we can call the broader bitcoin community (or 'bitcoiners'). Let this community consist of anyone who is sufficiently involved with bitcoin, such as by holding some bitcoin, or having done so in the past, or actively striving to hold some. It may be a tempting thought that we have to trust this community in light of the discussion so far: given that the features of bitcoin are due to what the bitcoin community collectively accepts, one could think that I need to trust that the bitcoin community continues to collectively accept the right things when I rely on bitcoin having the features that are due to these conventions.

The tempting thought that the bitcoin community needs to be trusted can be supported with an account of the collective acceptance involved in social construction, proposed by Passinsky. Passinsky proposes that collective acceptance involves taking on commitments: to collectively accept that *x* is F is to be committed to acting as if *x* is F (Passinsky 2020b: 437). According to this account, acceptance by a group comes with a commitment to act a certain way. Assuming also a commitment account of trust, there is then an argumentative path to the thought that engagement with bitcoin

requires trust in the bitcoin community to continue to act as if bitcoin has the relevant features.

This account cannot be quite right. Social construction of something by a community doesn't in general imply commitments on behalf of that community. Commitment implies intention and taking up a responsibility, but the sort of postulational conventions at stake in social construction may be unintentional and unconscious, and simply emerge from certain patterns of coordinated and intentional behaviour. Indeed, certain social entities are constructed but not intentionally so (Thomasson 2009: 549; compare Tuomela 2003: 129). Think of castes, economic recessions, housing markets or a public space. Social sciences are sometimes in the business of discovering social entities that are due to communities but not thereby also the results of intentional acts, nor already known. There is a natural distinction between the socially constructed entities that are intentionally created, and those that are unintentionally 'generated' by a community (Thomasson 2003). If this is right, mere social construction cannot imply a commitment, as this would imply that all social construction is intentional.

If social construction doesn't imply commitments to act a certain way, then either social construction doesn't involve collective acceptance, or (pace Passinsky) collective acceptance does not involve commitments to act a certain way. It seems to me that there can be collective acceptance that *x* is F without a commitment to act as if *x* is F. We should accept a weaker understanding of collective acceptance. One plausible candidate would be the view that there cannot be collective acceptance that *x* is F without a collective disposition to act as if x is F, where such a disposition may fall short of anything deserving to be called a commitment. This allows that collective acceptance of something can emerge unintentionally from the relevant social patterns.

Just as the community responsible for the social construction of bitcoin doesn't need to be trusted merely in engaging with bitcoin, the same could be said of the communities responsible for the social construction of ordinary money. If, in some Kafkaesque world, we stop overnight to be disposed to act as if paper bills are money, we thereby no longer collectively accept that paper bills are money and dropped the conventions that make those paper bills money. In this case too, there seems no breach of trust or a community failing to live up to a commitment; it would just be a social development about which one could be upset, in much the same way one can be upset about how the weather develops. This stands in contrast to the behaviour of the central banks, which do have a commitment that underwrites a level of trust in them doing what they are supposed to. Institutions can be founded with a certain intended purpose or a certain task, and the social fabric in which it is embedded may be such that this suffices for the institutions or groups to be committed to fulfilling this task. For example, one of the Federal Reserve's stated tasks is to keep inflation down and the economy stable, and this suffices for the Federal Reserve (or its members) to have a commitment to fulfilling this task.

What we can call the 'bitcoin community' is not some social institution founded to meet specified societal needs, it does not have a purpose, political, financial or otherwise, no central locus of decision making.

The defended view of the absence of commitments fit with existing discussions of the conditions under which there is group moral responsibility and group agency (Collins 2019; List and Pettit 2011, many thanks to an anonymous referee for pointing out this fit with existing discussions). The bitcoin community lacks the internal organization needed to bear responsibility and group agency, being no more than what Collins calls a 'combination' of people (Collins 2019: chap. 1).

If conventions regarding bitcoin change, one can be deeply disappointed or upset about this, like one can be upset about an expected change in the weather, but there would be no appropriate target for blame, there would be no group or no individuals one could feel appropriately betrayed by and any anger or protest would be misplaced. We may rely on bitcoin having and maintaining certain features, and hence indirectly rely on certain conventions and social patterns, but this needs to be carefully distinguished from relying on groups to fulfil certain commitments, and who need to be trusted.

There is a clear sense in which bitcoin is indeed aptly described as trust-free. When trust in the existing financial system is under pressure, the understandable response can be feelings of anger, betrayal, and indeed political upheaval, such as we saw in 2008 with the Occupy Wall Street protests. If the offered account is correct, this can never be the appropriate response to any developments in the code, nor in the social conventions that shape bitcoin.[2]

Leiden University, Netherlands.

References

Antonopoulos, A. (2017) *Mastering Bitcoin: Programming the Open Blockchain*. 2nd edition. Sebastopol, CA: O'Reilly Media.

Baier, A. (1986) 'Trust and Antitrust', *Ethics* 96/2: 231–60.

Bailey, A. M., B. Rettler, and C. Warmke. (2021a) 'Philosophy, Politics and Economics of Cryptocurrency I: Money Without State', *Philosophy Compass* 16/11: 1–15.

———. (2021b) 'Philosophy, Politics and Economics of Cryptocurrency II: The Moral Landscape of Monetary Design', *Philosophy Compass* 16/11: 1–15.

———. (forthcoming) *Resistance Money: A Philosophical Defense of Bitcoin*.

Bier, J. (2021) *The Blocksize War: The Battle over Who Controls Bitcoin's Protocol Rules*. Las Vegas, NV.

Collins, S. (2019) *Group Duties: Their Existence and Their Implications for Individuals*. New York: Oxford University Press.

[2] I'm grateful to two anonymous reviewers for their valuable feedback and to Jessica Brown for this invitation.

Fine, K. (2005) ‘Our Knowledge of Mathematical Objects’, in T. Z. Gendler and J. Hawthorne (eds.), *Oxford Studies in Epistemology*, 89–109. New York: OUP.
———. (2012) ‘Mathematics: Discovery or Invention?’, *Think* 11/32: 11–27.
Gladstein, A. (2022) *Check Your Financial Privilege: Inside the Global Bitcoin Revolution*. Nashville: BTC Media LLC.
Glazier, M. (2021) ‘Enterprise Blockchain Doesn’t Work Because It’s About the Real World’, Coindesk. <https://www.coindesk.com/markets/2021/03/31/enterprise-blockchain-doesnt-work-because-its-about-the-real-world/> accessed August 2022.
Hawley, K. (2014) ‘Trust, Distrust and Commitment’, *Noûs* 48/1: 1–20.
———. (2017a) ‘Applied Metaphysics’, in K. Lippert-Rasmussen, K. Brownlee, and D. Coady (eds.) *A Companion to Applied Metaphysics*, 165–79. New York: John Wiley & Sons.
———. (2017b) ‘Trustworthy Groups and Organisations’, in P. Faulkner and T. Simpson (eds.) *The Philosophy of Trust*, 230–250. New York: OUP.
———. (2018) ‘Social Science as a Guide to Social Metaphysics?’, *Journal for General Philosophy of Science* 49/2: 187–98.
———. (2019) *How to Be Trustworthy*. Oxford: OUP.
Jacobs, M. (2020) ‘How Implicit Assumptions on the Nature of Trust Shape the Understanding of the Blockchain Technology’. *Philosophy and Technology* 34/3: 573–87.
Kripke, S. A. (2011) ‘Vacuous Names and Fictional Entities’, in *Philosophical Troubles. Collected Papers Vol I*, 52–74. New York: OUP.
Lewis, D. (1986) *On the Plurality of Worlds*. New York: Wiley-Blackwell.
List, C., and P. Pettit. (2011) *Group Agency: The Possibility, Design, and Status of Corporate Agents*. New York: OUP.
Markosian, N. (2015) ‘The Right Stuff’, *Australasian Journal of Philosophy* 93/4: 665–87.
Mason, R. (2016) ‘The Metaphysics of Social Kinds’, *Philosophy Compass* 11/12: 841–50. https://doi.org/10.1111/phc3.12381.
Nakamoto, S. (2008) ‘Bitcoin: A Peer-to-Peer Electronic Cash System’, <http://bitcoin.org/bitcoin.pdf> accessed August 2022.
Passinsky, A. (2020a) ‘Should Bitcoin Be Classified as Money?’, *Journal of Social Ontology* 6/2: 281–92.
———. (2020b) ‘Social Objects, Response-Dependence, and Realism’, *Journal of the American Philosophical Association* 6/4: 431–43.
Quine, W. V. O. (1948) ‘On What There Is’, *Review of Metaphysics* 2/5: 21–38.
Searle, J. R. (1995) *The Construction of Social Reality*. New York: Free Press.
Smith, B., and J. Searle. (2003) ‘The Construction of Social Reality: An Exchange’, *American Journal of Economics and Sociology* 62/2: 285–309.
Thomasson, A. L. (2003) ‘Foundation for a Social Ontology’, *Proto-Sociology* 18: 269–90.
———. (2009) ‘Social Entities’, in R. Le Poidevin *et al.* (eds.) *The Routledge Companion to Metaphysics*, 545–554. Milton Park, UK: Routledge.

Tuomela, R. (2003) ‘Collective Acceptance, Social Institutions, and Social Reality’, *American Journal of Sociology and Economics* 62: 123–66.
Warmke, C. (2021) ‘What Is Bitcoin?’, *Inquiry: An Interdisciplinary Journal of Philosophy*. Online first.
———. (2022) ‘Electronic Coins’, *Cryptoeconomic Systems* 2/1: 1–28.