Electronic Coins[*]

Craig Warmke

Abstract

In the bitcoin whitepaper, Satoshi Nakamoto (2008: 2) defines an *electronic coin* as a chain of digital signatures. Many have since defined a *bitcoin* as a chain of digital signatures. This latter definition continues to appear in reports from central banks, advocacy centers, and governments, as well as in academic papers across the disciplines of law, economics, computer science, cryptography, management, and philosophy. Some have even used it to argue that what we now call bitcoin is not the *real* bitcoin. But the definition fails and Satoshi likely never endorsed it. In this paper, I explain why it fails and what Satoshi likely endorsed instead. Along the way, I untangle some issues around bitcoin fungibility and clarify some others around the ontology of digital assets.

1 Introduction

Bitcoin automates and decentralizes two tasks normally entrusted to third parties and centralized institutions. First, bitcoin's peer-to-peer network transfers value without trusted intermediaries. So anyone can send bitcoin directly to anyone without trusting payment processors and credit card companies. Second, no central bank controls the supply of bitcoin. Instead, nodes on the bitcoin network run software that encodes a fixed, disinflationary issuance schedule that caps the supply under 21 million around the year 2140.

Thus far, we've said something about how bitcoin *works.* But we've said very little about what bitcoins *are.* On this score, many have identified bitcoins with particular chunks of code. They draw inspiration from Satoshi Nakamoto, bitcoin's pseudonymous inventor. In the bitcoin whitepaper, Satoshi says: "We define an electronic coin as a chain of digital signatures."[1] An electronic coin's chain of digital signatures represents a transaction history similar to the way that a series of signatures does on a check. So, in the whitepaper, Satoshi defines an electronic coin as its encoded transaction history.

[*] [Removed]

[1] Nakamoto [2008, 2].

Many have since replaced the reference to electronic coins in Satoshi's definition and defined each *bitcoin* as a chain of digital signatures. We'll call this the *Chain Definition*. In effect, the Chain Definition identifies each bitcoin with an encoded transaction history. And it continues to appear in reports from central banks,[2] advocacy centers,[3] and governments,[4] as well as in academic papers across the disciplines of law,[5] computer science,[6] economics,[7] cryptography,[8] management,[9] and philosophy.[10] Some have even used it to argue that what we now call "bitcoin" is not the *real* bitcoin.[11] However, the Chain Definition fails, and, despite appearances, Satoshi likely never endorsed it.

Bitcoin is a highly interdisciplinary topic of research and often requires expertise in more than one field.[12] So we should sometimes expect researchers in one field to have skills or evidence relevant to a question that researchers in other fields lack. We find such an uneven distribution with respect to the status of the Chain Definition. Many bitcoin developers and computer scientists already know that the definition fails. But relatively few others do. And, in my experience, even the experts who do understand why the Chain Definition fails often lack the philosophical tools to understand Satoshi's original claim about electronic coins.

We can ask a number of philosophical questions about bitcoin, including some about bitcoin's relation to money.[13] But these questions come downstream from the more basic but equally philosophical question about what a bitcoin is in the first place. Here, I address this question about the nature of bitcoin by evaluating the most prominent answer. Overall, I have two aims. First, I hope to explain why the Chain Definition fails and thereby narrow the chasm between those closer to the epicenter of bitcoin development and those further away. Second, I hope to bring philosophical tools to bear on Satoshi's original claim about electronic coins. As we proceed, I will also untangle some related issues around bitcoin privacy and fungibility and clarify some others around the ontology of digital

[2] ECB [2012].

[3] Van Valkenburgh [2014, 9].

[4] Lastra and Allen [2018, 55].

[5] Akins et al. [2014, 30 n. 30], Zhang [2017, 560], Borroni [2016].

[6] Wu et al. [2017, 3124], Khalilov and Levi [2018].

[7] Bugár and Somogyvári [2020, 133].

[8] Gao et al. [2018, 27207].

[9] Friedlmaier et al. [2018, 2].

[10] Bjerg [2016, 3].

[11] Peter Rizun, at the Future of Bitcoin Conference, in 2017. See https://youtu.be/hO176mdSTG0?t=362 Those remarks occurred before the Bitcoin Cash hard fork, but Rizun repeats the criticism after the hard fork: https://twitter.com/PeterRizun/status/935285146562859008.

[12] Nathan Ballantyne [2019] calls questions *hybridized* when they require skills or evidence from different fields. For examples of hybridized questions about bitcoin, see Voshmgir and Zargham [ms.].

[13] Smit et al. [2016], Bjerg [2016].

assets more generally.

We begin in the next section by assessing the Chain Definition's meaning and motivation. In Section 3, I explain why the definition fails. Then, in Section 4, we return to Satoshi's definition of electronic coins. There, I explore the ontology of electronic coins and propose a framework for understanding Satoshi's definition. The paper concludes with some brief reflections on philosophy's role within the new interdisciplinary field of study concerning cryptoeconomic systems like bitcoin.

2 The Chain Definition

The Chain Definition identifies each bitcoin with a chain of digital signatures. But without some idea of what a bitcoin is, or what chains of digital signatures are, we'll be ill-equipped to evaluate the definition of one in terms of the other. So let's begin with a quick review of each.

2.1 Bitcoins

Bitcoin is a highly divisible asset, and its divisibility requires a unit of measurement. Somewhat confusingly, the word 'bitcoin' serves not only as a name for the asset (e.g., 'I have bitcoin'), but also as the unit of measurement for that asset (e.g., 'I sent 3.275 bitcoin') and a count noun for whole number amounts of the asset (e.g., 'He received four bitcoins'). In the count noun sense, the predicted maximum supply of bitcoin sits just under 21 million. And in this same sense, each bitcoin divides into 100 million units called *satoshis*. Hence, 2.37 bitcoin equals 237 million satoshis.

What does it mean to say that a *bitcoin* is a chain of digital signatures? As I'll soon explain, those who endorse the Chain Definition understand Satoshi's reference to "electronic coins" as a reference to bitcoins in the count noun sense. So they intend to identify chains of digital signatures with the things we expect someday to number under 21 million and that neatly divide into 100 million satoshis.

Although I've not seen anyone define a satoshi as a chain of digital signatures, there's perhaps more reason to think Satoshi's definition of electronic coins concerns satoshis rather than bitcoins. First, by definition, every bitcoin transaction transfers a whole number amount of satoshis. Second, bitcoin transactions are actually denominated in satoshis within the *blockchain*, the public ledger of bitcoin transactions. Wallets, exchanges, and other services have converted the raw transaction amounts denominated in satoshis to the presently more readable amounts denominated in bitcoin.[14] But instead of evaluating whether satoshis are chains of digital signatures, we will continue to focus on the Chain Definition, which

[14] Antonopoulos [2017, 121-122].

concerns bitcoins and remains popular. The main arguments below apply equally well to both claims.

2.2 Digital Signatures

What does it mean to say that a bitcoin is a *chain of digital signatures*? And, more fundamentally, what is a digital signature? Digital signatures in bitcoin work like physical signatures on checks. So we may better understood digital signatures by analogy with their physical counterparts.

When Alice writes a physical check to Bob, she specifies an amount to credit Bob's account from her own and signs the check to authorize the transaction. When Bob deposits the check, the bank then performs two tasks. First, the bank verifies Alice's signature. By verifying signatures, banks help ensure that no one but the owner(s) of an account transfers funds from it. Second, after the bank ensures that Alice's account has sufficient funds, the bank clears the transaction by debiting Alice's account and crediting Bob's with the specified amount. So the bank serves as a trusted intermediary by verifying signatures and clearing transactions. The bitcoin network verifies signatures and clears transactions without trusted intermediaries.

Let's first review how the bitcoin network disintermediates the verification process. We will simplify matters here and abstract away from some unnecessary details. Suppose Alice has received a bitcoin at one of her addresses and would like to send it to Bob. She then uses a software application to compose a candidate transaction that sends that bitcoin to one of Bob's addresses. To compose such a transaction, Alice must include some important pieces of information:

(i) the *public key* which generates her address through a mathematical function.

(ii) the previous transaction output(s) that credited the address in (i) with the bitcoin yet to be spent.

(iii) one or more receiving addresses and the amount to send each. The total spending amount cannot exceed the amount in (ii), and here Alice specifies an amount of 1 bitcoin for one of Bob's addresses.

Not just anyone can spend Alice's bitcoin, though, even if someone provides all the above information. For Alice's transaction to appear in the ledger, she needs the address's private key, a string of characters that functions like the address's password. The private key bears a mathematical connection to its address: the public key in (i) which generates the address through a mathematical function is itself generated by the private key through another mathematical function. And this private key enables

Alice to produce a digital signature that is otherwise practically impossible to provide.

How does Alice's private key help produce the digital signature? The signature results from feeding a special digital signature function two chunks of information. Roughly, the first chunk is the information embedded in (ii) and (iii). The second chunk is Alice's private key.[15] The diagram below shows how a private key generates the public key which, in turn, generates the address, and also how the same private key helps produce the digital signature for a transaction that spends bitcoin previously received at the private key's address. The middle column represents the transaction:

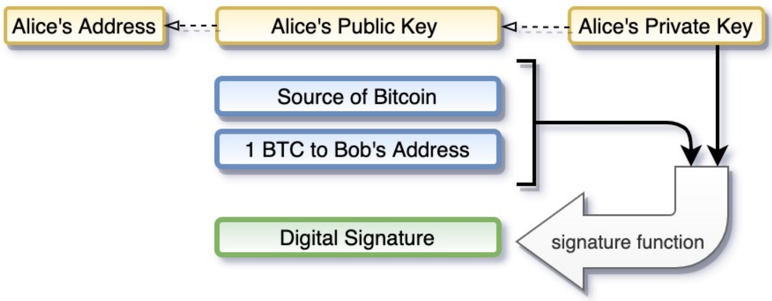


Figure 1. *Alice composes a transaction of 1 BTC to Bob's address.*

The resulting signature is important for two reasons. First, producing the signature without the private key is practically impossible. Second, anyone can easily verify that Alice's private key helped produce the signature without access to the private key. To verify a signature over a proposed transaction, a verification function only needs the signature and the proposed transaction (in Figure 1, the information in the middle column), all publicly available information. Computers running the bitcoin software called *full nodes* use this function to verify signatures and reject candidate transactions whose signatures fail to verify. Full nodes also reject attempts to spend previously spent bitcoin, as well as other kinds of ill-formed attempts.

In addition to disintermediating the verification of signatures, the bitcoin network also disintermediates the clearing of transactions. Full nodes send valid transactions to *miners*, computers running the bitcoin software that compete to publish those transactions in the ledger's next block.[16] About every ten minutes, a miner wins a race to solve a well-defined mathematical puzzle and sends a block with the solution back to the full nodes to verify. Once the full nodes verify the block and append it to their own

[15] See Antonopoulos [2017, 139 ff.] for more details.

[16] I'll also use 'miners' to refer to the people who use their capital to mine bitcoin.

individually stored copy of the ledger, the race begins anew. Miners compete in this way because the winning block rewards the miner with new bitcoin out of thin air in the amount determined by the disinflationary issuance schedule. But if miners use more or less processing power and begin to solve puzzles on average either more or less quickly, the puzzle difficulty automatically adjusts so that the solution time remains close to ten minutes.

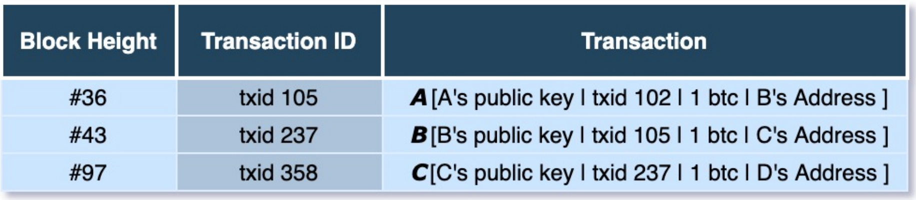
However, transactions can be undone, in principle. Blocks of transactions have a degree of clearance that depends on how many blocks have been published since. Due to the way blocks are ordered, undoing a transaction—and its block—requires creating an alternative chain that branches off from a previous block. The alternative chain then grows with additional blocks from that point forward. But bitcoin's consensus mechanism endorses whichever chain furnishes evidence of having used the most processing power to solve puzzles. Therefore, undoing a transaction would require creating an alternative chain with evidence of having used more total processing power than the original, even as the original continues to grow. Hence, the further back a block appears in the blockchain, the safer it is because one would have to solve puzzles at a faster pace than the miners incentivized to work on the original chain. And if miners try but fail to create such an alternative chain, they lose all the mining rewards that might have covered the costs of creating it. So the mining reward lures people to compete in a way that secures the value of that very reward. But because someone could theoretically use enough power to undo a past block, transactions are cleared only in a probabilistic sense.

2.3 The Definition's Appeal

Now that we've covered how the bitcoin network automates the verification and clearance functions, we can appreciate the appeal of the Chain Definition. Although the table below abstracts away from crucial details involved in transactions, it includes enough detail to trace a bitcoin's trajectory through a chain of signatures across the blockchain. The simplified transactions in the table below should be read as having the following form:

Sender Signature [sender public key | sender source of bitcoin | trans. amount | receiver]

Transactions are, of course, much more complicated than this. But we can still recognize the Chain Definition's appeal at such a high level of abstraction.



Block Height	Transaction ID	Transaction
#36	txid 105	***A*** [A's public key l txid 102 l 1 btc l B's Address]
#43	txid 237	***B*** [B's public key l txid 105 l 1 btc l C's Address]
#97	txid 358	***C***[C's public key l txid 237 l 1 btc l D's Address]

Table 1. *Tracing a bitcoin's history through digital signatures.*

In the table's last transaction, D's address receives a bitcoin. Where did it come from? Well, this last tranasction bears C's digital signature, which enabled C to spend the bitcoin sent to C's address in txid 237. Then, txid 237 bears B's digital signature, which enabled B to spend the bitcoin sent to B's address in txid 105. And txid 105 bears A's signature, which enabled A to spend the bitcoin sent to A's address in txid 102. We've now traced a chain of digital signatures through a bitcoin's transaction history.

A typical transaction in bitcoin's ledger specifies one or more sources of bitcoin to spend and one or more destinations for that bitcoin to be spent. But unlike a ledger whose entries represent beers owed at the bar or gold owned in a vault, the entries on the bitcoin ledger represent nothing "out in the world." So if bitcoins aren't outside the ledger, perhaps they are somehow embedded in the ledger itself. So perhaps each bitcoin just is the chain of digital signatures in the ledger that represents its transaction history.

Yet not just any transaction histories will do. In 2012, the European Central Bank published a report on virtual currencies. After quoting Satoshi's definition of electronic coins as chains of digital signatures, the report claims that the very bitcoins which are "divisible to eight decimal places" are such that each "carries the entire history of the transactions it has undergone, and any transfer from one owner to another becomes part of the code."[17] The European Central Bank has correctly inferred that if bitcoins are transaction histories, they are entire transaction histories. For identifying them with partial histories would imply that there have been more bitcoins than the system says there have been. But what is an entire transaction history?

All bitcoin first appears in a *coinbase transaction*, the special transaction that rewards a winning miner's address with newly minted bitcoin. And all bitcoin remains unspent at the address to which it was most recently spent. So if each bitcoin has an entire transaction history, that history should consist of a chronologically ordered series of transactions from the originating coinbase transaction through all and only the subsequent transactions in which that bitcoin is signed over to an address. For any given bitcoin, then, if a series of transactions excludes a transaction in

[17] ECB [2012, 23].

which that bitcoin is signed over, the series isn't that bitcoin's *entire* history. And if a series of transactions includes a transaction in which that bitcoin isn't signed over, the series isn't *that* bitcoin's entire history. Therefore, *if* bitcoins are chains of digital signatures, they are chains of digital signatures that represent entire transaction histories.

We may now offer an official formulation:

Chain Definition. A bitcoin is a chain of digital signatures that represents its entire transaction history.

Many have endorsed either the above definition, something that implies the above definition, or something near enough. Now, some might wonder whether those who espouse the Chain Definition use 'bitcoin' to mean something other than what we ordinarily mean by it today. But the surrounding context often makes clear that authors mean to identify chains of digital signatures with bitcoins as we think of them today. We've already mentioned the report from the European Central Bank. But let's run through a few more examples. In computer science, Wu et al. [2017, 3124] say of the very bitcoins which will someday total near 21 million that "Nakamoto defines them as chains of digital signatures." Similarly, and closer to my own field of philosophy, Bjerg [2016, 3-5] claims of the very bitcoins which will someday number under 21 million that each "consists of a unique chain of digital signatures." In law, after specifying that 'bitcoin' refers to the unit of account, Zhang [2017, 556, 560] says that "each electronic bitcoin consists of a 'chain' of 'digital signatures'..." In economics, Kroll et al. [2013, 1-6] offer a slightly weaker version of the definition. They say that the very "Bitcoins" which were valued at $130 at the time of the paper's writing, and the very things whose number halves every four years in the mining reward, are such that each is "represented as a chain of digital signatures over the transactions in which the coin was used." We could provide more examples across more disciplines. But these will suffice for now.

The Chain Definition has achieved enough interdisciplinary influence to merit a quick but public execution. Though we shouldn't have inferred that bitcoins are chains of digital signatures from Satoshi's definition of electronic coins as digital signatures, the mistake is quite understandable. In fact, I conclude in Section 5 that Satoshi's own writings and the nature of interdisciplinary research made this mistake nearly inevitable. But, first, let's see why bitcoins are not chains of digital signatures.

3 Why the Chain Definition Fails

To see how the Chain Definition fails, we'll follow three stages of bitcoin transactions. When we look at these transactions closely, we see that they transfer quantities of bitcoin and not individuals with entire transaction histories, as the Chain Definition implies. And once we see why the Chain Definition fails, we can better understand the clever engineering decision that underlies the varying levels of fungibility in the bitcoin system.

3.1 Main Argument

Stage One. In this first stage, addresses 1 and 2 (A1 and A2) each send one bitcoin to the previously empty A3:

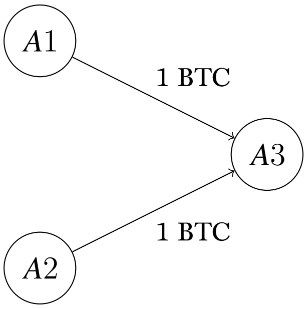


Figure 2. Two bitcoins from different addresses to the same address.

The two bitcoins at A3 are distinguishable and not perfectly fungible. But what distinguishes them? They don't have different names or identifiers because bitcoins don't have names or identifiers. Nor do satoshis. Instead, every quantity of unspent bitcoin is uniquely tied to its most recent transaction. To see how each bitcoin at A3 is uniquely tied to its most recent transaction, we must dive deeper into the technical details of transactions.

Bitcoin transactions have both inputs and outputs. An output in a transaction specifies an amount of bitcoin and an address as the recipient of that amount. Transactions have one or more outputs, and each has an index number within its transaction. The first or only output in any transaction is 0, the second is 1, and so on. Consequently, we can identify any output in the blockchain by referring to its index number and associated transaction ID. Transactions also have one or more inputs, and each input tags as its source of bitcoin an output from a previous transaction by its index number and transaction ID. However, within transactions, outputs do not tag particular inputs as their sources of bitcoin. Even so, the total amount across a transaction's outputs cannot exceed the total amount across its inputs. When we put it all together, each transaction specifies, in its inputs, the bitcoin to spend by referencing outputs from previous transactions, and, in its outputs, how much of that bitcoin goes where.[18]

The bitcoins at A3 are distinguishable because they were sent to A3 in different transactions. In the diagram below, we see how each quantity of one bitcoin is tied to a different output from a different transaction:

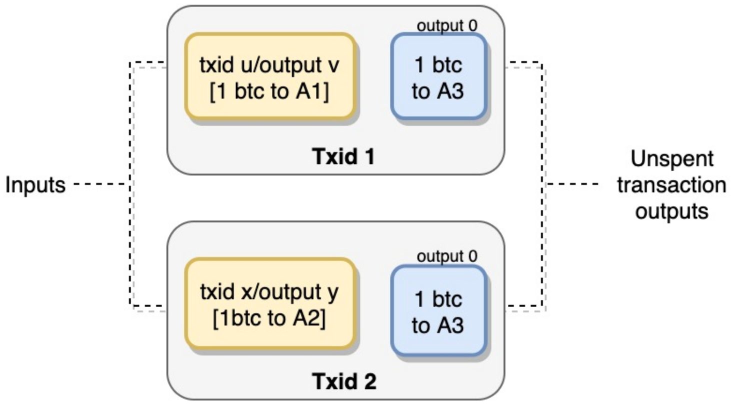


Figure 3. The inputs and outputs from the two transactions in Figure 2.

Because the bitcoins at A3 are distinguishable, the possessor of A3's private key may spend the bitcoin from A1 rather than the bitcoin from A2 (and vice versa). To spend the bitcoin from A1, the user would specify the transaction output in which A1 sent A3 a bitcoin as the source of bitcoin in a new transaction input. Alternatively, to spend the bitcoin from A2, the user would specify the transaction output in which A2 sent A3 a bitcoin as the source of bitcoin in a new transasction output.

Stage Two. Instead of spending one bitcoin rather than another, A3 then sends both to A4 in a single transaction:

[18] In this way, bitcoin uses a "transaction-based" rather than an "account-based" ledger, the kind banks ordinarily use. Hal Finney [2008] makes the distinction and categorizes bitcoin correctly two weeks after the whitepaper's publication. For an accessible explanation of the distinction, see Akcora et al. [2018, 2-3]. Although the choice between these two kinds of ledgers has important tradeoffs, they differ less than some let on. For a translation scheme between the two kinds of ledgers, see Zahnentferner [2018]. Because of translation schemes like this, it is wholly appropriate to liken transaction inputs in the ledger to debits and transaction outputs in the ledger to credits as Antonopoulos [2017, 18] does.

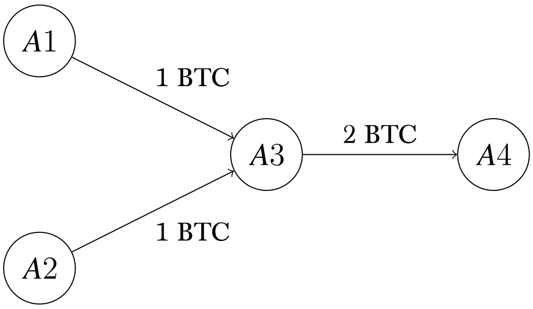


Figure 4. A single transaction with two bitcoins from one address to another.

Previously, in Stage 1, the bitcoins at A3 were distinguishable because they arrived at A3 in different transactions. In Stage 2, however, both bitcoins arrive at an address in a single transaction. So the kind of feature which previously distinguished the bitcoins at A3 no longer distinguishes them at A4. Nothing else distinguishes them either. No names, no tags—nothing. The bitcoins are now perfectly fungible with one another, and it simply isn't possible to tag one rather than the other to spend in a new transaction. We see the same when we look at the transaction's two inputs and single output:

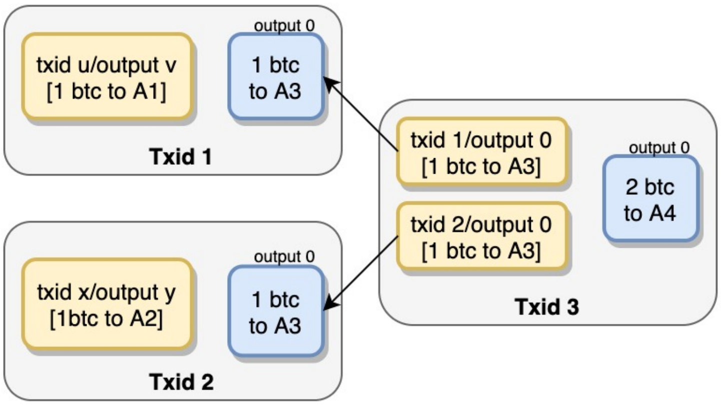

txid u/output v [1 btc to A1]
output 0
1 btc to A3
Txid 1
txid x/output y [1btc to A2]
output 0
1 btc to A3
Txid 2
txid 1/output 0 [1 btc to A3]
txid 2/output 0 [1 btc to A3]
output 0
2 btc to A4
Txid 3

Figure 5. The inputs and outputs from transactions in Figure 4.

The single output of two bitcoin to A4 does not distinguish the two bitcoins in any way. It simply represents a primitive quantity of two bitcoin. Of course, the owner of A4 could then spend an amount of one bitcoin out of two.[19] But the owner cannot single out and spend an individual bitcoin because there are no individual bitcoins to be singled out and spent.

[19] In bitcoin transactions, one spends the entirety of the previously unspent bitcoin from a previous transaction. But one can send a specified amount back to the same address (which is discouraged for security reasons) or to another address for which one has the private key.

There are no individual bitcoins within the code or represented by the code. There are only represented quantities of bitcoin.

The point is not that bitcoins are fictional and that fictions lack identity criteria. For the sake of argument, let's grant that bitcoin is fictional and that some fictional entities like Sherlock Holmes have identities within their fictional universes. The point is that the "two bitcoins" in A4 are not differentiated entities, even within any supposed fiction.[20] At most, we have a quantity of a fictional substance that does not decompose into distinguishable individuals.

An analogy may help. Suppose you deposit $1 from your friend and $1 from your sibling in your previously empty savings account. Asking which bitcoin in A4 came from A1 and which came from A2 is like asking which of the two dollars came from your friend and which came from your sibling. The question falsely presupposes that digital dollars carry this information on the bank's ledger. Similarly, the Chain Definition falsely presupposes that there are individual bitcoins that carry individuating features on the bitcoin ledger. Just as there is no fact of the matter about the source of any individual digital dollar in your account, there is no fact of the matter about the source of any individual bitcoin in A4. In both cases, we have quantities without individuals.

Stage Three. It gets worse for the Chain Definition. Not only are there no individual bitcoins to pair with entire transaction histories, there is often no entire transaction history to pair with a purported individual bitcoin. To bring this point into relief, we will proceed to the next stage of transactions. In this third stage, A4 sends one bitcoin to each of A5 and A6:

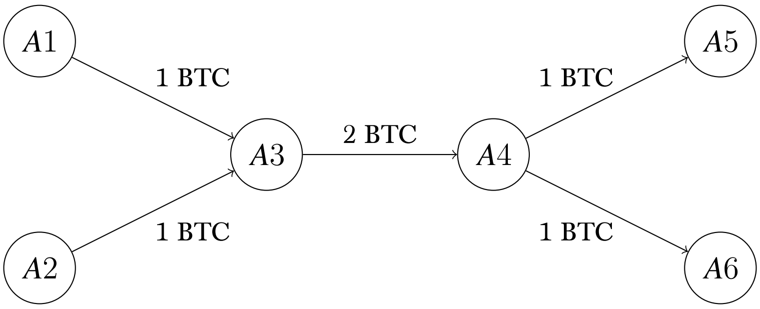


Figure 6. A transaction which halves an amount of 2 BTC across two addresses.

As we covered in Stage 2, there are not two individual bitcoins in A4 but only a quantity of two bitcoin. As a result, there is no fact of the matter about *which* bitcoin arrives in A5 rather than A6. So there is no fact of the matter about whether the bitcoin in A1 ultimately arrives at A5 or A6.

[20] For orthogonal reasons, I do argue in [Removed] that bitcoin is a fictional substance.

Likewise, there is no fact of the matter about whether the bitcoin in A2 ultimately arrives at A5 or A6. We see this more clearly when we zoom in on the transaction details. The single transaction output new in Stage 2 gets claimed in the single transaction input new in Stage 3. The new transaction in Stage 3 splits a primitive quantity into two smaller primitive quantities:

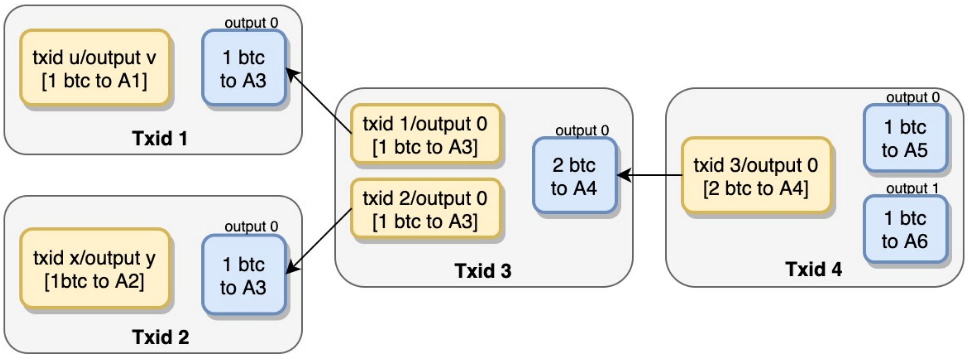


Figure 7. The inputs and outputs from transactions in Figure 2.

Now, if the bitcoin ledger did mark bitcoins as individuals, we could discover whether the bitcoin at A5 took the A1-A3-A4-A5 path or the A2-A3-A4-A5 path. But the ledger doesn't mark bitcoins as individuals. So when we try to trace the history of the bitcoin at A5, the path "goes dark" in txid 3. The blockchain encodes no information about whether the bitcoin at A5 was involved in the transaction through A1 rather than the transaction through A2. Similar remarks apply to the bitcoin in A6.

In the idiom of the bitcoin community, this series of transactions has *mixed* the bitcoin initially sitting separately at A1 and A2. Mixing produces a metaphysical and not merely epistemic indeterminacy. We don't merely not know which path a bitcoin has taken. There is no such path. As a result, neither the bitcoin in A5 nor the bitcoin in A6 has an entire transaction history. Hence, neither bitcoin has an entire transaction history with which it could be identical.

Mixing is important not only because it falsifies the Chain Definition. It also enhances bitcoin fungibility and financial privacy. The features that distinguish quantities of bitcoin from each other pre-mixing get smeared across the quantities of bitcoin post-mixing. In the Stages above, mixing occurs in two transactions by funneling separate quantities of bitcoin into a single output, then a single input, and out again through multiple outputs. But, often, users cooperate to mix quantities of bitcoin within a single transaction. By jointly constructing a multiple input and multiple output transaction, users can send bitcoin back to themselves at new addresses with smeared histories. This CoinJoin method, developed by Gregory Maxwell [2013], has been advertised as enhancing privacy because it severs the connection between users and their past addresses. But it also enhances fungibility by smearing the sources of transaction inputs across all the outputs.

Some might think that in multi-input/multi-output transactions like CoinJoins, inputs pair up explicitly with outputs as if to say "Input 2 is the source of bitcoin in Output 3." But bitcoin transactions don't work this way. We see where a transaction's bitcoin comes from and where various amounts go, but no links exist between a transaction's particular inputs and outputs. For this reason, CoinJoin sometimes draws comparisons to money laundering. But quantities of bitcoin undergo similar kinds of mixing and smearing on a regular basis as a part of the network's normal function. Mixing occurs not only within multi-input/multi-output transactions but also over patterns of transactions like we saw in Stages 1 through 3.

3.2 Assessment

The Chain Definition fails because it falsely presupposes that the bitcoin ledger marks bitcoins as individuals. But without individual bitcoins, we don't have entire transaction histories either. What could a bitcoin's entire transaction history be except for the path that a specific bitcoin takes through a series of transactions? In general, a history of an individual piggy-backs ontologically on that individual. No individual, no history. So if there are no individual bitcoins, there are no entire transaction histories either. Consequently, the Chain Definition faces double jeopardy. The Chain Definition falsely presupposes that the bitcoin ledger marks bitcoins as individuals and also identifies bitcoins with transaction histories that fail to exist precisely because the ledger does not mark bitcoins as individuals.

Now, perhaps the Chain Definition goes slightly too far in identifying bitcoins with entire transaction histories. As I mentioned in Section 2.3, Kroll et al. [2013] say that "a bitcoin is a fixed-value cryptographic object *represented* as a chain of digital signatures over the transactions in which the coin was used."[21] Something can *represent* something else without being *identical* to it. However, the very features of bitcoin transactions which sink the Chain Definition also sink the claim that each bitcoin is "represented as a chain of digital signatures over the transactions in which the coin was used." Without names or markers for individual units, the blockchain does not encode enough of the right information in transactions to distinguish those units. Neither does it encode the sort of transaction histories whose existence requires that missing information. Mixing patterns like those depicted in Figures 2 through 7 illustrate this point well enough.

[21] The emphasis is mine.

With the exception of coinbase transactions, the bitcoin protocol ensures that a transaction's outputs preserve the total amount of bitcoin from its inputs, much like the laws of thermodynamics ensure that, in an isolated system, energy is not created or destroyed. But, in preserving these amounts, the network uses transactions that transfer quantities of bitcoin, not individuals that are bitcoins. In so doing, the bitcoin network resolves what we may call the *divisibility dilemma*. To ensure that no smallest unit gets spent more than once at a time, one might have thought that each such unit needs an identity on the ledger. But, then, if users were to transact in highly divisible amounts, and each unit required a digital signature to spend, transactions could easily require many millions of signatures. With much use, the network would eventually buckle.[22] So a system that tracks the identities of individual units faces a trade-off between scalability and divisibility. Satoshi avoided this trade-off altogether by adopting a ledger of the kind described by Wei Dai [1998] that tracks primitive quantities instead of individual units with identities. So the Chain Definition ultimately obscures one of Satoshi's smarter engineering decisions.

4 Instruments and Quantities

If bitcoins are not chains of digital signatures, how should we understand Satoshi's definition of electronic coins as chains of digital signatures? The definition did not occur in a vacuum. Bitcoin built on previous attempts at electronic cash, and these attempts help illuminate the reference to electronic coins in the bitcoin whitepaper.

A substantial number of proposals for electronic money preceded bitcoin's arrival. These proposals naturally inherited features from their non-electronic counterparts. To illustrate, consider a signed check for $20. The check is a *financial instrument* whose scribblings signify a quantity of $20. And, except in special cases, what signifies isn't identical to what's signified. A note with the name 'Einstein' signifies the man, Einstein. And neither the note nor the name is the man. The same generally holds for financial instruments and the quantities they signify. In the case of our $20 check, the signifying instrument, its scribblings, and the signified quantity have different properties. Whereas the check is an individual and bears a traceable identifier on its face, the quantity of $20 itself has no traceable identifier. Different properties, different identities.

One might object that the signified quantity does have a traceable identifier: it inherits the identifier of the signifying check. But the connection between the quantity and the identifier is contingent, and we can easily sever it. We can deposit the check in a previously empty bank account

[22] As Tatsuaki Okamoto [1995, 439] correctly observes, "a system in which a coin worth $5367 consists of 5367 $1 coins is a rather unwieldy and inefficient divisible cash system." Compare Chow [2007, 151].

and then tear up the check. The quantity survives in the bank account as the check and its identifier sit in the garbage bin.

The distinction between signifying instrument and signified quantity does not disappear when we digitize the instrument. For example, in the 1980s, David Chaum created the eCash system for transferring electronic money in the form of Cyberbucks.[23] In this system, Alice creates a random string of symbols to serve as a numeric note analogous to a paper check with a check number. Then, she sends the note to a bank for a digital signature that determines the note's monetary value. If she wanted a note for 20 Cyberbucks, for instance, the bank would withdraw that amount from her account and sign the numeric note with the private key reserved specifically for signing 20-Cyberbuck notes.

Like checks and their signified quantities, each note differed from the quantity of Cyberbucks signified. The notes were strings of symbols and had identifiers in the form of their randomly chosen note numbers. They were digital financial instruments. But whereas the digitized instrument bore a traceable identifier in the form of its note number, the signified quantity of Cyberbucks had no such traceable identifier. Now, any particular quantity of Cyberbucks might have had a contingent tie to the note that signified it. But these ties were contingent and easily severed. We could deposit the note in a previously unused bank account where the quantity of Cyberbucks would persist but the numeric note and its identifier would not.

The distinction between the signifying financial instrument, on the one hand, and the signified quantity, on the other, holds for bitcoin's other predecessors, too. For example, the RPOW tokens developed by Hal Finney [2004] were financial instruments in the form of a string of symbols. Whereas physical tokens signify a quantity with numerals stamped on their sides, Finney's RPOW tokens signified a quantity with symbols encoded in their bits. Unfortunately, Finney didn't use a special unit of account like the Cyberbuck. So let's call the unit an *rpow*. Then, we can say that Finney's RPOW tokens signified quantities of rpow. Again, quantities of rpow were not themselves the RPOW tokens, not even if an RPOW token signified a lone rpow.

The distinction between instrument and quantity also holds for bitcoin. But if the blockchain represents quantities of bitcoin, what are the signifying financial instruments? The financial instruments are unspent transaction outputs or *UTXOs*. UTXOs are transaction outputs that remain unspent. They are like physical checks that have yet to be signed and deposited. Unsurprisingly, UTXOs and their signified quantities of bitcoin differ in important ways. Whereas each UTXO has an identifier in the form of its index number and the ID of the transaction in which it appears, no quantity of bitcoin has such an identifier. And while spending

[23] Chaum [1983, 1985, 1992], Chaum et al. [1988].

a UTXO "destroys" it (in the sense that the network no longer considers it unspent), the signified quantity of bitcoin may survive as the signified quantity of a new UTXO.

By distinguishing signifying instrument and signified quantity, we may begin to diagnose the Chain Definition by asking two questions. First, what are the "electronic coins" in the whitepaper? And, secondly, what does Satoshi mean when he defines an electronic coin as a chain of digital signatures? In relation to the first question, we might expect Satoshi's "electronic coins" to be the digitized instruments, like Chaum's notes and Finney's RPOW tokens. Then, we could charge the Chain Definition with substituting a reference to the the digitized instruments we call UTXOs with a reference to individual bitcoins. Since the blockchain does not represent bitcoins as individuals, and since digitized instruments differ from the quantities of bitcoin they signify, we would expect such a substitution to fail. If only things were so simple.

In many discussions about bitcoin, 'unspent transaction output' is ambiguous between the chunk of code that signifies a quantity of unspent bitcoin and the signified quantity itself. Let's reserve 'UTXO' for the signifying chunk of code. And let's call the signified quantity of bitcoin an *unspent quantity*. We can then eliminate UTXOs as the focus of Satoshi's original claim. Although both UTXOs and chains of digital signatures are chunks of code, they are very different chunks of code. Whereas chains of digital signatures include bits of code from different transactions, each UTXO is embedded as an output within a single transaction. So it wouldn't make much sense to define one in terms of the other.

We can make better sense of Satoshi's definition if we understand the "electronic coins" in the whitepaper as unspent quantities of bitcoin. Early bitcoin users would often transfer a UTXO's entire unspent quantity without leaking any in transaction fees, like pouring a cup's contents into another without spilling. Feeless transactions occur rarely now, so a UTXO's unspent quantity is now typically a flash in the pan. Even so, each UTXO represents an unspent quantity which has persisted at least vacuously through a chain of one or more digital signature. So we can meaningfully pair a UTXO's signified quantity of unspent bitcoin with the chain of digital signatures over the transactions which have preserved that quantity.

On the question about the meaning of Satoshi's definition, the most charitable interpretation would say that Satoshi doesn't mean to identify chains of digital signatures with quantities of unspent bitcoin. Chains of digital signatures and unspent quantities of bitcoin have different properties and cannot be identical. While the former are chunks of code the latter are not. Identifying Sherlock Holmes with the name 'Sherlock' is involves one kind of confusion. Identifying Sherlock with the instances of punctuation that end the sentences including 'Sherlock' involves a more needlessly complex sort of confusion. This latter identification closely parallels the absurd idea that unspent quantities of bitcoin are chains of digital signatures.

Instead of identifying unspent quantities of bitcoin with chains of digital signatures, Satoshi more likely offers a model of the kind we often find in math, philosophy, science, and so on. Theorists often use sets of objects to model the meanings of words, sets of possible worlds to model propositions (and vice versa), *n*-tuples of real numbers to model locations in *n*-dimensional real space, and sets themselves to model numbers. Many who model in this way do not thereby identify the things that model with the things being modeled. And we *can* model a UTXO's particular quantity of unspent bitcoin with the chain of digital signatures that has preserved that quantity back to the transaction in which it resulted by combining smaller quantities, by splitting a bigger quantity, or by serving as a mining reward. Therefore, we can paraphrase Satoshi as saying something like this: "Let a chain of digital signatures represent an unspent quantity of bitcoin." It makes sense to represent unspent quantities of bitcoin with certain chains of digital signatures, and it makes much more sense to represent *them* with chains of digital signatures than it does to represent individual bitcoins or UTXOs with chains of digital signatures.

However, we should note that if this interpretation is correct, then Satoshi departs from tradition by using a word for a financial instrument to refer to the unspent quantities the relevant instruments signify. 'Coin', like 'note' and 'token', typically refers to an instrument and not the quantity the instrument signifies. And, as we saw above, Satoshi's predecessors followed this tradition in discussing their digitized instruments. But this doesn't seem to me a serious objection against the proposal, since Satoshi clearly departs from this way of speaking in another way. For example, a little later on, Satoshi uses 'coins' for neither the instruments nor their quantities but instead for the units themselves.[24] Admittedly, this makes it easier to read the earlier use of "electronic coin" in the whitepaper as a reference to the units, too. But, as I've argued, we should resist this temptation and instead let charity, and the surrounding context, guide our interpretation.

5 Conclusion

In the bitcoin whitepaper, Satoshi says that "electronic coins" are chains of digital signatures. Many have since endorsed the Chain Definition by inferring from Satoshi's claim that bitcoins are chains of digital signatures. But the inference fails. As I argued in Section 3, the Chain Definition falsely presupposes that the bitcoin ledger marks bitcoins as individuals.

[24] Nakamoto [2009].

The ledger tracks quantities of bitcoin, not individual bitcoins. But if the electronic coins in Satoshi's definition are not bitcoins, what are they? Many would say that they are UTXOs. But we must distinguish UTXOs from the quantities of bitcoin that they signify. And I've argued that the electronic coins in the whitepaper are probably best understood as these quantities of bitcoin and not the UTXOs that signify them.

Unfortunately, the Chain Definition has begun to spread like an interdisciplinary wildfire. Like many wildfires, the Definition's influence has been both destructive but understandable. It has been destructive not only because it has sown confusion across many disciplines, but also because some used it as a key premise to defend the contentious bitcoin cash hard fork in 2017. Nonetheless, the definition's influence has been understandable since Satoshi uses "coin" equivocally very early on. For example, as early as the v0.1 Bitcoin software release, Satoshi refers to the 21,000,000 maximum supply of bitcoins as a cap on "coins."[25] This equivocal use of terminology made some confusions almost unavoidable, especially when so many academics with interests in bitcoin operate at an altitude far above its technical machinery.

Bitcoin sits at the core of a new and highly interdisciplinary field of study where more confusions await us. This field of cryptoeconomics will not succeed without legal scholars, mathematicians, economists, and computer scientists speaking across disciplinary boundaries.[26] But this very condition for success will draw some to trespass into disciplines for which they have little or no training.[27] Some will gatekeep from an understandable desire to protect their turf. Many others will talk past each other with superficially similar terminology. And, of course, some will attempt to exploit the confusion for personal gain. Going forward, those who specialize in clarifying concepts and drawing distinctions could play an invaluable role.[28]

[25] Nakamoto [2009].

[26] Voshmgir and Zargham [ms.].

[27] Ballantyne [2019].

[28] Though I have philosophers foremost in mind, see Walch [2016, 2017, 2019] for examples from a legal perspective.

References

Virtual currency schemes, 2012. European Central Bank. Frankfurt am Main.

Cuneyt Gurcan Akcora, Matthew F. Dixon, Yulia R. Gel, and Murat Kantarcioglu. Blockchain data analytics. *Intelligent Informatics*, page 4, 2018.

Benjamin W Akins, Jennifer L Chapman, and Jason M Gordon. A whole new world: Income tax considerations of the bitcoin economy. *Pitt. Tax Rev.*, 12:25–56, 2014.

Andreas Antonopoulos. *Mastering Bitcoin: unlocking digital cryptocurrencies.* O'Reilly Media, Inc., 2nd. edition, 2017. First published in 2014.

Nathan Ballantyne. Epistemic trespassing. *Mind*, 128:367–395, 2019.

Ole Bjerg. How is bitcoin money? *Theory, Culture & Society*, 33(1):53–72, 2016.

Andrea Borroni. Bitcoins: Regulatory patterns. *Banking & Finance Law Review*, 32(1):47, 2016.

Gyöngyi Bugár and Márta Somogyvári. Bitcoin: Digital illusion or a currency of the future? *Financial and Economic Review*, 19(1):132–153, 2020.

David Chaum. Blind signatures for untraceable payments. In *Advances in cryptology*, pages 199–203. Springer, 1983.

David Chaum. Security without identification: Transaction systems to make big brother obsolete. *Communications of the ACM*, 28(10):1030–1044, 1985.

David Chaum. Achieving electronic privacy. *Scientific american*, 267(2): 96–101, 1992.

David Chaum, Amos Fiat, and Moni Naor. Untraceable electronic cash. In *Conference on the Theory and Application of Cryptography*, pages 319–327. Springer, 1988.

Sherman SM Chow. Running on karma–p2p reputation and currency systems. In *International Conference on Cryptology and Network Security*, pages 146–158. Springer, 2007.

Wei Dai. b-money, 1998. URL http://www.weidai.com/bmoney.txt.

Hal Finney. Rpow: Reusable proofs of work, 2004. URL https://nakamotoinstitute.org/finney/rpow/.

Hal Finney. Bitcoin p2p e-cash paper, 2008. URL https://www.metzdowd.com/pipermail/cryptography/2008-November/014848.html.

Maximilian Friedlmaier, Andranik Tumasjan, and Isabell M Welpe. Disrupting industries with blockchain: The industry, venture capital funding, and regional distribution of blockchain ventures. In *Venture Capital Funding, and Regional Distribution of Blockchain Ventures (September 22, 2017). Proceedings of the 51st Annual Hawaii International Conference on System Sciences (HICSS)*, 2018.

Yu-Long Gao, Xiu-Bo Chen, Yu-Ling Chen, Ying Sun, Xin-Xin Niu, and Yi-Xian Yang. A secure cryptocurrency scheme based on post-quantum blockchain. *IEEE Access*, 6:27205–27213, 2018.

Merve Can Kus Khalilov and Albert Levi. A survey on anonymity and privacy in bitcoin-like digital cash systems. *IEEE Communications Surveys & Tutorials*, 20(3):2543–2585, 2018.

Joshua A Kroll, Ian C Davey, and Edward W Felten. The economics of bitcoin mining, or bitcoin in the presence of adversaries. In *Proceedings of WEIS*, volume 2013, 2013.

Rosa Maria Lastra and Jason Grant Allen. Virtual currencies in the eurosystem: Challenges ahead. *Brussels, Belgium: ECON Committee, European Parliament*, 2018.

Gregory Maxwell. Coinjoin: Bitcoin privacy for the real world, 2013. URL https://bitcointalk.org/?topic=279249.

Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008. URL http://bitcoin.org/bitcoin.pdf.

Satoshi Nakamoto. Bitcoin v0.1 released, 2009. URL https://satoshi.nakamotoinstitute.org/emails/cryptography/16/#selection-9.0-9.21.

Arvind Narayanan and Jeremy Clark. Bitcoin's academic pedigree. *Communications of the ACM*, 60(12):36–45, 2017.

Tatsuaki Okamoto. An efficient divisible electronic cash scheme. In *Annual International Cryptology Conference*, pages 438–451. Springer, 1995.

Kalle Rosenbaum. *Grokking Bitcoin*. Manning, Shelter Island, 2019.

JP Smit, Filip Buekens, and Stan Du Plessis. Cigarettes, dollars and bitcoins–an essay on the ontology of money. *Journal of Institutional Economics*, 2(2):327–347, 2016.

Peter Van Valkenburgh. Comments to the conference of state bank supervisors on the draft model state regulatory framework for virtual currency. 2014.

Shermin Voshmgir and Michael Zargham. Foundations of cryptoeconomic systems. ms.

Angela Walch. The path of the blockchain lexicon (and the law). *Review of Banking & Financial Law*, 36:713–765, 2016.

Angela Walch. Blockchain's treacherous vocabulary: One more challenge for regulators. *Journal of Internet Law*, 21(2):9–16, 2017.

Angela Walch. Deconstructing 'decentralization': Exploring the core claim of crypto systems. In *Crypto Assets: Legal and Monetary Perspectives*. Oxford University Press, 2019.

Qianhong Wu, Xiuwen Zhou, Bo Qin, Jiankun Hu, Jianwei Liu, and Yong Ding. Secure joint bitcoin trading with partially blind fuzzy signatures. *Soft Computing*, 21:3123–3134, 2017.

Joachim Zahnentferner. Chimeric ledgers: Translating and unifying utxo-based and account-based cryptocurrencies. *IACR Cryptology ePrint Archive*, 2018:262, 2018.

Yilu Zhang. The incompatibility of bitcoin's strong decentralization ideology and its growth as a scalable currency. *NYUJL & Liberty*, 11:556, 2017.