The Economics of Bitcoin Mining, or Bitcoin in the Presence of Adversaries

Joshua A. Kroll, Ian C. Davey, and Edward W. Felten
Princeton University

Abstract

The Bitcoin digital currency depends for its correctness and stability on a combination of cryptography, distributed algorithms, and incentive-driven behavior. We examine Bitcoin as a consensus game and determine that it relies on separate consensus about the rules and about game state. An important aspect of Bitcoin's design is the *mining* mechanism, in which participants expend resources on solving computational puzzles in order to collect rewards. This mechanism purportedly protects Bitcoin against certain technical problems such as inconsistencies in the system's distributed log data structure. We consider the economics of Bitcoin mining, and whether the Bitcoin protocol can survive attacks, assuming that participants behave according to their incentives. We show that there is a Nash equilibrium in which all players behave consistently with Bitcoin's reference implementation, along with infinitely many equilibria in which they behave otherwise. We also show how a motivated adversary might be able to disrupt the Bitcoin system and "crash" the currency. Finally, we argue that Bitcoin will require the emergence of governance structures, contrary to the commonly held view in the Bitcoin community that the currency is ungovernable.

1 Introduction

Bitcoin [7, 22] is a decentralized electronic fiat currency implemented using cryptography and peer-to-peer technology. At the time of writing, there are over 11.2 million Bitcoins in circulation which can be traded for a wide variety of goods and services and for which liquid exchange markets exist for at least 18 other currencies. Bitcoins trade in volatile exchange markets; recent prices have fluctuated around $115/BTC (historically, $5-10/BTC was common; at the time of writing, prices were closer to $130/BTC), meaning that the Bitcoin monetary base is currently just over one billion dollars. While Bitcoin and other cryptographic digital currencies are typically analyzed for security properties (e.g. no double-spending[1] of coins), their construction and protocols are rarely

[1] *Double spending* occurs when a digital coin is duplicated and spent more than once. Virtual coins are easily duplicated, so a digital currency must have some defense against it.

analyzed additionally for their *economic* soundness. Only by pairing a careful technical analysis with the relevant economic factors can we determine whether the Bitcoin protocol is stable. In this paper, we examine the *stability* of Bitcoin from an economic and technical perspective.

Like any fiat currency, Bitcoins have value by consensus and by virtue of the ability to use them to purchase goods and services. But Bitcoin is more than just a currency; it is also a distributed algorithm which must function correctly in order for the currency to operate, for example to maintain a consensus as to who owns which coins. The successful operation of these algorithms relies in turn on assumptions that participants in the system will cooperate in certain ways. Whether it is safe to assume cooperation depends on whether the parties' incentives induce them to cooperate.

Ultimately, Bitcoin relies on three types of consensus. Participants must maintain consensus (1) on the rules to determine validity of transactions, (2) on which transactions have occurred in the system, and (3) that the currency has value. The three forms of consensus are connected, in the sense that the failure of any one will unravel the other two.

In this paper we address two main questions. First, we ask whether the Bitcoin protocol is stable, in the sense that the system will continue to operate if all parties act according to their incentives.

Second, we ask whether a malicious participant, who wants to disrupt Bitcoin and destroy its value, will be able to do so. In particular, we consider a new class of attack: the *Goldfinger attack*,[2] in which the attacker's motivation is based on some incentive outside the Bitcoin economy. Such an adversary might, for example, be a law enforcement or intelligence agency which wishes to see Bitcoin holdings weakened.[3] Equally, an adversary might have significant short positions in Bitcoin exchange markets. Or, as suggested by Becker et al. [6], such an adversary might be distributed in the form of a social protest movement opposed to activity in the Bitcoin community.

Finally, we consider how threats to the Bitcoin community, in the form of actual adversaries, protocol instabilities, and inevitable bugs and accidents, necessarily require mechanisms for governance. We argue that such governance is already emerging, that it will take the form of the governance of an open source project (in the sense that leaders cannot take actions contrary to the interests and will of the community without naturally losing legitimacy), and that the emergence of formal governance structures will ultimately subject Bitcoin itself (and not merely particular players) to influence by government regulators around the world.

[2] Auric Goldfinger is the villain in an eponymous 1964 film [15]. Goldfinger wanted to increase the value of his own gold holdings by making the gold in Ft. Knox radioactive and thus worth less. Many aspects of Bitcoin are described by analogy to the gold standard, making the comparison especially apt.

[3] It is known that Bitcoins are used to facilitate a significant trade in illegal goods on Silk Road, an anonymous online marketplace that has over $8 million in monthly sales [5].

Structure of the Paper

In Section 2, we explain how the Bitcoin protocol works. In Section 3, we model Bitcoin mining, the core of the Bitcoin protocol, as a game played by miners and Bitcoin holders. Section 4 explores the equilibria of this game and examines the effects of the “51% attack”, in which a cartel of miners dictates outcomes in the game. In Section 4.2, we discuss the transaction fee mechanism and its problems. In Section 5, we examine how the mining game might change in the presence of a Goldfinger-type adversary. In Section 6, we discuss the emergence and necessity of governance in Bitcoin. Section 7 concludes by arguing that, contrary to the claims of the Bitcoin community, Bitcoin will naturally fall under the (limited) sway of government regulation over time.

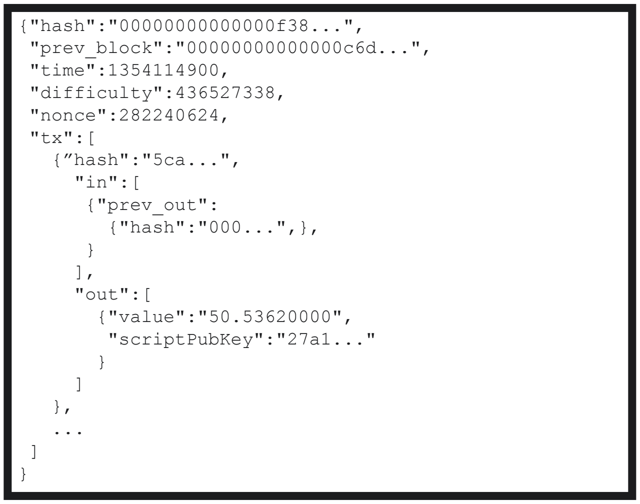
2 Background: How Bitcoin Works

Bitcoin is a cryptographic currency based on ideas from Hashcash [3] and b-money [11] which aims to be completely distributed, free of central authorities or points of control, and at least somewhat anonymous. Rather than a detailed written specification, Bitcoin is defined by a short white paper published under a pseudonym [22], together with a reference implementation [7]. In practice, questions about the rules are answered primarily by inspecting the behavior of the reference implementation. We will describe anyone who holds a Bitcoin or participates in the Bitcoin peer-to-peer network as a Bitcoin *player*.

A Bitcoin is a fixed-value cryptographic object represented as a chain of digital signatures over the transactions in which the coin was used. A coin can be checked for validity simply by checking the cryptographic validity of the signatures that constitute its history. Each Bitcoin is owned by a *Bitcoin address*, which consists of a public key.[4] The owner of a Bitcoin (that is, the holder of the corresponding private key) can create a transaction (acting as the *sender*) by signing an assertion that Bitcoins are being transferred from one address to another. A transaction may involve many input identities and many output identities. Occasionally an extra output value will appear in a transaction for *change* to transfer back to the sender, since fixed-value coins must be transferred in an all-or-nothing manner. If the total value of the input Bitcoins exceeds the value of the output Bitcoins, the difference is interpreted as a *transaction fee*, which is paid to the player who successfully appends that transaction to the *block chain*, a globally-consistent log data structure which is described below.

Although the above protocol allows the receiver of a Bitcoin transaction to verify cryptographically that the transaction is a valid payment order, it does not prevent *double spending* of Bitcoins. That is, while the receiver can verify that the sender did at one point own the Bitcoins being transferred, he has no

[4] Technically, addresses in Bitcoin are comprised of a cryptographic hash of the public portion of an ECDSA key pair and a check sum. But since Bitcoins can also be sent directly to an ECDSA public key, we will ignore this subtlety.



```
{"hash":"00000000000000f38...",
 "prev_block":"00000000000000c6d...",
 "time":1354114900,
 "difficulty":436527338,
 "nonce":282240624,
 "tx":[
   {”hash":"5ca...",
     "in":[
      {"prev_out":
        {"hash":"000...",},
      }
     ],
     "out":[
       {"value":"50.53620000",
        "scriptPubKey":"27a1..."
       }
     ]
   },
   ...
 ]
}
```

Figure 1: A schematic block in the block chain, represented as JSON (JavaScript Object Notation). Some metadata has been left out and truncated values are marked by ellipses.

way to know if the coins he is being given have been previously used to pay someone else.

To prevent double spending, Bitcoin players engage in a peer-to-peer protocol that implements a distributed timestamp service providing a fully-serialized log of every Bitcoin transaction ever made. Transactions are organized in the log into *blocks*, which contain a sequence number, a timestamp, the cryptographic hash of the previous block, some metadata, a *nonce*, and a set of valid Bitcoin transactions. A schematic representation of an individual block is shown in Figure 1. The blocks form a hash chain: each new block contains the cryptographic hash of its predecessor, allowing anyone to verify that no preceding block has been modified. The block chain contains backward links but not forward links (a block cannot link forward to a future block that has not yet been created) so there is a unique path backward from each block to the beginning of the log (the *genesis block*) but the forward path from a block might not be unique. Thus the log has the form of a tree whose branches fork as it grows. The block chain is shown in Figure 2.

Any player may choose to become a *miner* and *mine* new blocks that add new transactions to the log. A new block is a valid addition to the log if its nonce is chosen so that the new block's hash is less than a *target* value. This is a form of proof-of-work puzzle, a computation that is thought to be difficult to perform but whose result is easy to verify. The solution to a proof-of-work puzzle effectively asserts that someone has expended a certain level of effort [12, 17, 6].

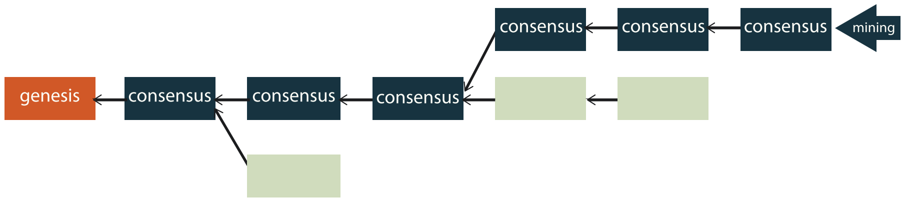


Figure 2: An example abstract blockchain. The genesis (first) block is on the left. Mining occurs on the longest branch of the branching tree. Other branches and branches with invalid blocks are ignored.

The specific proof-of-work in Bitcoin is taken from Hashcash [3]. The difficulty of the proof-of-work puzzle is adjusted periodically by an adaptive algorithm based on the recent block chain history to maintain the long-term invariant that one new block be mined every ten minutes on average.

The mining mechanism has the property that if there are two branches of the tree, with a separate group of miners growing each branch, then the branch whose miners have more computational power will grow more quickly[5]. In a sense, miners vote for a branch by devoting their mining effort to extending it, and the Bitcoin rules say that the longest branch should be treated as the only valid one.

When a user Alice wishes to transfer Bitcoins to Bob, she creates and signs a transaction object and broadcasts it to her peers in the Bitcoin peer-to-peer network. The peers then rebroadcast it, effectively flooding the network with all known pending transactions.[6] All of the miners (that is, players who also elect to mine) then attempt to create a new block with the pending transactions they know about.

New Bitcoins can only be created via the mining process. Each miner adds to their prospective block a special transaction creating a number of *reward* Bitcoins which may be paid to anyone (but which typically are paid to the miner). This provides an incentive for miners to engage in mining. The number of Bitcoins created this way is adjusted on a predetermined schedule in which the reward is halved each time 210000 more blocks have been mined. The original

[5] Because miners search randomly for puzzle solutions, a branch supported by fewer mining resources might happen to grow faster in the short run, but in the long run the branch with more resources will always win. Prudent Bitcoin participants will wait for a while before accepting one branch as valid, to eliminate the possibility that the longest branch is short-term lucky and will lose in the long run. Karame, Androulaki, and Capkun [16] describe attacks that are possible if participants do not wait.

[6] But see Babaioff et al. [2], which posits that if a transaction has a transaction fee, this transaction flooding itself might not be incentive-compatible. In Section 4.2, we examine the issue of incentives from transaction fees from a different perspective, arguing that such fees might not be a reasonable basis for the mining game.

mining reward was 50 Bitcoins (50 BTC) per block, but was cut to 25 BTC when block 210000 was mined in November 2012. Mining does not guarantee a reward; the first miner who happens to find a suitable solution will extend the block chain and claim the mining reward. Then all miners start over, trying to solve a new puzzle to add yet another block to the block chain.

Bitcoins are generally considered anonymous because Bitcoin addresses are derived from public keys and so could represent anyone on the Internet. In practice participants might be identifiable. Transactions made under multiple identities, such as payments to oneself, can be linked in some circumstances [26], and transaction behavior can leak identifying information [1]. Finally, it should be noted that while Bitcoin grants a high level of anonymity when deployed by a user herself, the bar for correct deployment is quite high. Many users keep their Bitcoins on deposit with a large exchange such as Mt. Gox [21], which thus have a great deal of information identifying Bitcoin holders. Recently suggested extensions to the Bitcoin protocol provide provable and strong anonymity guarantees, but are not yet deployed [20].

Bitcoin comes at the end of a long and rich history of digital currency efforts going back over 30 years. As early as 1982, Chaum had articulated many of the key ideas in implementing electronic cash [10]. This leaves the natural question of why Bitcoin has been so successful when other systems have failed to catch on, as addressed recently by Barber et al. [4].

3 Basic Mining Economics

3.1 Consensus

Success of the Bitcoin economy requires that Bitcoin's distributed protocols operate and remain stable. In this section we consider the stability of these protocols, under the assumption that players behave according to their incentives.

The success of Bitcoin relies on three types of consensus:

- Consensus about Rules: Players must agree on criteria to determine which transactions are valid. Only valid transactions will be memorialized in the Bitcoin log; but this requires agreement on how to determine validity.
- Consensus about State: Players must agree on which transactions have actually occurred, that is, they must agree on the history of the Bitcoin economy, so that there is a common understanding of who owns which coin at any given time.
- Consensus that Bitcoins are Valuable: Players must agree that Bitcoins have value so that players will be willing to accept Bitcoins in payment.

Each of these forms of consensus depends mutually on the other two. For example, it is hard to agree on the history without agreeing on the rules. And it is hard to believe in the value of a Bitcoin if participants cannot even agree on who owns which Bitcoin.

Consensus about the rules is a social process. Participants must come to a common understanding of what is allowed, so that the rules can be encoded into the software that each participant uses. In Bitcoin, small groups and individuals can exert outsized power.[7]

Consensus about state is a technological problem in distributed systems design. Each player can see part of the state and the players need to cooperate, in large numbers and across a potentially unreliable network, to achieve a consistent understanding of the global state. Technological consensus must be achieved despite the possibility that some players will deviate from the published rules. In the distributed systems literature, devious behavior ("Byzantine failures") can often be tolerated if a sufficient majority of players are honest and cooperate. However, in Bitcoin, we explicitly assume that players will behave according to their incentives. (Assuming cooperation despite incentives to the contrary would make the design much simpler, though unrealistic.)

Finally, consensus that Bitcoins are valuable is the same sort of consensus necessary for any fiat currency. Such value is often modeled as a focal point in a coordination game (because players need something to use as a medium of exchange and a unit of account, they choose a local currency because it is available). Such an analysis is necessary but not sufficient to explain the consensus that Bitcoins are valuable.

3.2 Modeling the Mining Process

A simple model illuminates how players decide whether and how to mine. Imagine a new Bitcoin player, Minnie, who wishes to determine whether or not to become a miner. Minnie has the option to invest resources (say, equipment and electricity) in mining at a cost of C dollars per second, and must decide whether to make the investment. Say that this investment will allow Minnie to make $P = f(C)$ puzzle guesses (hashes) per second, a puzzle takes G guesses to solve in expectation, and that successfully solving a puzzle (i.e., mining a block) gives a reward of Bitcoins with value V. Finally, assume that all players face the same decision (i.e. that no player has access to special technology or significant discounts that are not available to other players[8]). Then Minnie will earn PV/G expected dollars per second and so will invest if

$$G < \frac{PV}{C}$$

However, Minnie is not the only miner. Recall that the number of guesses to solve a puzzle G is dependent on the rate of mining recently so that the overall

[7] An example is *mining pools*, which are collaborative mining efforts used to smooth the mining payout. Some mining pools are thought to represent over 30% of the total mining capacity. Additionally, it is known that there are a few concentrated holdings of Bitcoin which are each in excess of 1% of the total supply, such as the widely published holdings of Tyler and Cameron Winkelevoss [25].

[8] Note that this means that our model does not capture *casual* (that is, non-professional) mining, such as mining by botnets or on corporate machines. Such tactics are used, for example, to externalize the cost of electricity and equipment acquisition and maintenance.

global rate of mining new blocks is held constant, say R new blocks per second. Assume there are N miners globally. Then at any time,

$$R = \sum_{i=1}^{N} \frac{P_i}{G}$$

Let $\bar{P}$ be the total number of hashes (guesses) per second globally, the numerator of the above expression, and $\bar{C} = \sum_{i=1}^{N} C_i$ be the total spent on mining globally. Then $G = \bar{P}/R$ and so Minnie will want to enter the mining market only if:

$$\frac{\bar{P}}{R} < \frac{\bar{P}V}{\bar{C}} \Longrightarrow \bar{C} < RV$$

Thus, we would expect to have a global equilibrium in which the total mining reward in dollars per second is equal to the total global cost of mining:

$$\bar{C} = RV$$

This equilibrium will hold as long as miners can make an instantaneous decision about whether or not to mine. In practice, miners are amortizing sunk costs related to capital expenditure for equipment, meaning $\bar{C}$ has a fixed component and a marginal component. Miners may therefore overinvest in mining to offset their fixed costs, investing up to their marginal costs per unit time.

In Bitcoin circles, the total number of hashes per second made by all players is referred to as the *network hash rate*. At the time of writing, the network hash rate was estimated to be approximately 119 trillion hashes per second (119 Thash/s). This makes the Bitcoin network one of the largest distributed computing projects ever undertaken: taken as a whole, the Bitcoin transaction verification network is more powerful than the combined computing power of the top 500 supercomputers in the world, giving pause to anyone concerned about whether the costs of transaction verification in Bitcoin are acceptable [6]. However, the hash rate cannot be measured directly because the Bitcoin log only contains solutions to puzzles and not an accounting of how much computation was required to find those solutions. Several projects attempt to model the hash rate on an ongoing basis [30, 24].

This mining equilibrium leads us to an interesting conclusion about Bitcoin: because mining resources must currently be purchased with currencies other than Bitcoin, the value of the mining reward V fluctuates with the exchange price of Bitcoin. Thus, if the Bitcoin price falls substantially, so too does the incentive to mine. This leads to the possibility of a *death spiral* in which loss of confidence in Bitcoin could cause the Bitcoin price to go down, a falling price lowers the incentive to mine and the equilibrium mining rate, lower mining rate leads to the currency being easier to subvert, and this leads to a further loss of confidence in the currency. Such a death spiral reflects the perceived loss of consensus in the value game; we observe that it can happen even when consensus in the other games is functioning (as in an exchange rate crash) but that loss of consensus for the rules or for the game state contribute directly to the loss of consensus for value. We examine this further in Sections 4 and 5.

Our analysis suggests that it is important that mining activity does not generate any extra value for the miner (beyond the mining reward itself). If Bitcoin were changed so that a unit of mining effort with cost C generated inherent value $f < C$ to the miner (e.g., by computing answers to valuable problems), then the effort expended by miners would increase by a factor of $\frac{C}{f}$ in equilibrium, so that the same amount of resources would be wasted as in the current case. We do note, though, that it would be useful to change the mining process so that it created value that could not be captured by the miner, for example by attacking a problem whose solution would be a pure public good.

3.3 Mining Strategy

Now imagine that Minnie has decided to become a miner, investing C^* to buy $P^* = f(C^*)$ hashes per second. Minnie now has to make choices about *how* to mine. The Bitcoin documentation states certain rules that Minnie is supposedly required to follow, but we assume here that Minnie will act to maximize her utility, regardless of what might be written in the Bitcoin documents.

None of the rules of Bitcoin are self-executing; any rule can be ignored by users. Consider, for example, the rule requiring that a transaction carry valid digital signatures from the owner of every input coin. Everyone can use cryptography to *detect* a violation of the rule. But the rule will only be *enforced* if players ignore transactions that do not carry a cryptographically valid signature. Cryptographic rules and other technical rules are like all other rules, in that they exist only as words on paper and therefore will be followed only to the extent that players have incentives to follow them.

How should Minnie mine to ensure the maximum expected return? The main decision she faces is where in the Bitcoin log structure she should try to construct new blocks. Although the documents often speak of the Bitcoin log as a “chain” of blocks, in general the log could fork, perhaps at several points, leading to a structure that is more like a branching tree than a single linear sequence of blocks. In principle, Minnie’s mining effort, which aims to create a new block, could be aimed to extend any of the existing branches or to create a new branch anywhere in the tree.

The Bitcoin documents say that miners are supposed to try to extend the longest branch, but this rule is only words on paper. If miners all follow this rule, then the longest branch will tend to grow even longer, and in case of a fork one branch will soon outrun the other. But do miners have an incentive to follow the longest-branch rule, or do they have an incentive to behave otherwise, for example to create or sustain forks? Forks are thought to be dangerous to Bitcoin because they create multiple, competing versions of the transaction history and thus sow doubt about who owns which coins.

We will model mining as a game played by all miners. Each miner chooses a *strategy* S. A strategy is a function that maps the block chain structure L (up to the current round) to a choice of which branch to mine on (that is, which block will be the parent of the newly-mined block if the player wins). That is, $S(L) = b^*$ means that S selects b^* when given the log L. Each player chooses a strategy before playing. The payoff for each miner is their expected return from the mining reward for the block mined in each round. However, that reward is only valuable if the newly-mined block ends up on the long-term consensus chain.[9]

We call a strategy S *monotonic* in the Bitcoin history if, for two block chain structures L_r and L_{r+1} that differ only by the addition of a new block b with parent block $S(L_r)$, it is the case that $S(L_{r+1}) = b$. The intuition is that if the strategy is trying to extend the tree from a particular point, then the addition of a new block at that point causes the strategy to move on to the newly added block. The longest-chain strategy, as specified in the Bitcoin documents, is monotonic—if the longest chain is extended by one block, it remains the longest chain. However, there are infinitely many monotonic strategies.

We observe that if S is a monotonic strategy, then the mining game has a Nash equilibrium in which all miners play S. Let us consider Minnie's choice of mining strategy S^* in a game in which all other players have committed to playing a monotonic strategy S. If Minnie plays S, then all players will be playing the same monotonic strategy, and monotonicity implies that a single branch will grow without bound. As a result, every block that Minnie successfully mines will be on the long-term consensus branch. If Minnie were to switch to another strategy, she could not create new blocks any faster, because her rate of block creation is independent of which branch she is on. The only effect of switching strategies would be to make it possible for some or all of the blocks she creates not to be on the long-term consensus branch. Thus deviating from strategy S can only lower her utility, proving that $(S, S, ..., S)$ is a Nash equilibrium.

Our model resembles the *stag hunt* or *trust assurance* game in the literature [29]. However, in the typical stag hunt, both mutual cooperation and mutual defection are equilibrium outcomes. In Bitcoin, we observe that mutual defection, while possibly stable, will not be acceptable to the players, who will take steps to restore the cooperating equilibrium, even going as far as to change the protocol rules. We discuss the interdependent types of consensus required for Bitcoin to operate in Section 3.1.

The above analysis shows that any monotonic strategy is a Nash equilibrium when adopted by all players. In this sense, no monotonic strategy is better than any other. Why then, in practice, do all players choose to follow the longest-chain strategy if other strategies would also lead to equilibria? New players will follow the strategy chosen by a majority of the existing miners. But how does the majority arrive at this strategy in the first place? We suggest that the choice of strategy is a tacit coordination problem [27] and thus the solution which seems most attractive serves as a *focal point*. Miners chose the longest-chain strategy initially because it was used in the reference Bitcoin implementation. As new mining capacity has entered the system, that choice has proved stable.

[9] A block is "on the long-term consensus chain" if, in the limit as time goes to infinity, that block will be on the longest path in the tree with probability one. In some cases there may be no blocks, or only a bounded number of blocks, on the long-term consensus chain.

We consider below in Section 4 whether a sufficiently motivated player could disrupt this stability.

4 The Stability of the Mining Game

In this section, we look at several scenarios relevant to the *stability* of the equilibria described above. That is, we examine whether the system will return to equilibrium if perturbed.

In this section and throughout the paper, we consider attacks which aim to destabilize consensus about the rules or state of Bitcoin. Obviously, other classes of manipulation exist such as classical currency manipulation (for example, standard pump-and-dump schemes). We are interested, however, in the stability of the Bitcoin game and so will not consider these valid attack scenarios.

4.1 A Cartel of Miners, or the 51% Attack

The security of Bitcoin relies on the distributed consensus achieved by the mining game. In our analysis thus far, we have assumed, as the Bitcoin developers do, that a *cartel* of miners cannot form. That is, no coordinating group of miners (or a single player) can hold more than 50% of the network's mining (puzzle-solving) capacity. However, this assumption is questionable: mining is now generally organized into *pools* of coordinating miners who partition the search for proof-of-work puzzle solutions and who share in the mining rewards. One such pool, BTC Guild [9], controls over one quarter of total mining power. Furthermore, as mining becomes increasingly specialized (with specialized hardware such as application-specific integrated circuits (ASICs) for efficient hashing and the need for powerful computers to validate transaction blocks), the barriers to entry increase, effectively concentrating the mining power among a few powerful players who are less accountable to the (much larger) set of all Bitcoin holders. We naturally ask, therefore, what a mining cartel could do if one ever comes to exist.

First, we observe that a cartel can change any rules which are enforced by consensus and players who are not in the cartel will likely be obliged to follow. For example, a cartel can choose any strategy in the mining game. Players who continue to use the old strategy risk having their newly-mined blocks ignored as forks of the consensus branch and thereby risk losing the mining reward payments associated to those blocks. Thus, if the cartel announces its mining strategy, it can shift the equilibrium chosen by the non-cartel players.

It is often asserted (for example, in the Bitcoin white paper [22]) that a cartel can double-spend Bitcoins. In a strict sense, this is true: a cartel can spend a Bitcoin by paying it to a player Alice, receiving goods or services, and then shifting the consensus choice of history to a branch where that coin is instead paid to a different player Bob. However, we argue that double-spending by a cartel has a limited payoff. Bitcoins have value because people are willing to trade them for goods and services. If players were unwilling to accept

Bitcoins for trade or unwilling to spend Bitcoins for fear of having their payments nullified, the value of Bitcoins would diminish significantly as players lost confidence in the system. Worse, because players are encouraged to generate a new identity for each transaction and because identities are not linked to any side information, players cannot easily determine whether a proffered payment is coming from the double-spending cartel or an honest user. Thus, a rational player should refuse to accept *any* payments when there is a significant threat of double-spending. As a cartel must outmine the entire Bitcoin network and thus outspend the entire Bitcoin network for as long as it would remain a cartel, we believe it is very unlikely that a cartel could double-spend enough to recover the cost of the attack.

An interesting facet of mining cartels is that they can censor certain transactions. The cartel can choose to ignore any transaction it does not want appended to the log. Further, the cartel can choose to treat any blocks appended by others to the log as forks which it will not attempt to extend. Thus, other players will naturally also abandon these transactions, possibly even consciously if the cartel announces that certain transactions (or transactors) are disfavored.

4.2 Transaction Fees

The Bitcoin protocol allows a transaction to leave a “transaction fee” for the miner. If the value paid out of a transaction (in Bitcoins) is less than the amount put in, the difference is treated as a transaction fee that can be collected by whoever manages to mine a block containing that transaction. A transaction fee is like a tip or gratuity left for the miner.

A miner’s incentive is to include in their mined block any transaction that offers a nonzero transaction fee. All else being equal, the miner is better off accepting even a tiny transaction fee, rather than passing up the fee by refusing to mine the transaction. Although the miner might wish the transaction fee were larger, every miner knows that if they refuse to process a transaction, another miner can process it and collect the fee. If miners try to make an agreement to boycott users who leave small transaction fees, the agreement will not hold—individual miners will be able to defect from the agreement, by mining under anonymous identities. Such an agreement can be modeled as a prisoner’s dilemma: the miners would benefit from cooperation but cannot agree not to defect; users may wish to leave a tiny transaction fee to make sure the miners will process a transaction, but the user gets no benefit from offering a larger fee—doing so simply gives up value without any compensating gain.

As a result, we expect an equilibrium in which users leave small, nonzero transaction fees, and miners collect those fees. Indeed, this is what we observe: the expected total mining reward sans fees in a day is 3600 BTC, while the average total daily transaction fee take is only about 50 BTC.[10] The transaction

[10] We should note, however, that the reference Bitcoin client leaves a small, nonzero transaction fee by default. Users may override this default and opt to send transactions without fees. Many (but not all) users do exactly this, but the default may lead to higher-than-equilibrium fees.

fee mechanism is similar to the classic ultimatum game [23], and thus one might initially think that a norm for leaving nonzero transaction fees could evolve. We do not believe this is the case: a Bitcoin player proffering a very small fee transaction is not giving an ultimatum to a *specific* miner, but rather offering any successful miner in the future an opportunity to collect the fee by including the transaction in the log. As long as any miner is willing to accept such transactions, fees will be bid down.

We therefore do not expect transaction fees to play a significant long-term role in the economics of the Bitcoin system, under the current rules. We believe that a rules change would be necessary before transactions fees can play any major role in the Bitcoin economy. We discuss this possibility below in Section 6.2.

5 The Goldfinger Attack

As described above, a 51% cartel attack is unlikely to generate enough reward within the Bitcoin economy to be worthwhile to the attacker. However, this does not rule out the possibility of a 51% attack that aims to destroy the Bitcoin economy in order to achieve utility *outside* the Bitcoin economy. We call this the *Goldfinger attack* after the character in film who tries to undermine U.S. currency by ruining its gold backing [15].

There are at least three possible motivations for a Goldfinger attack. First, a government or institution might want to block Bitcoin transactions, to enforce the law, deter money laundering, or achieve some other institutional goal. Second, a non-state attacker might seek to gain some political or social goal, perhaps as a form of social protest (such a model was previously postulated by Becker et al. under the name "Occupy Bitcoin" [6]). Third, an attacker might seek an investment gain, for example by taking large short positions in Bitcoins so as to profit if the value of Bitcoins is diminished.

In all of these cases, the attacker must achieve enough utility to justify the substantial cost of an attack. We agree with Becker et al. that it is unlikely that a protest movement could muster the resources to launch a successful attack. And at present it does not appear possible to acquire a short position on Bitcoins that is large enough to justify an attack.

It follows that governments are the most plausible source of Goldfinger attacks, perhaps as a law enforcement tactic. Bitcoin is used to manage a significant trade in illicit goods such as in the anonymous online market Silk Road [5]. Traditional law enforcement techniques do not function well in the context of Bitcoin, which lacks a central issuing authority or direct, useful measures for tracking down the identity of players behind specific transactions. Nonetheless, there is significant law enforcement interest in shutting down these illegal activities: the FBI has even issued a report on the use of Bitcoin for criminal activity [13] and two sitting U.S. senators have sent a letter to the Department of Justice and the Drug Enforcement Administration urging them to crack down [28].

5.1 Modeling Goldfinger Attacks

We model the possibility of a Goldfinger attack by a simple game with two players: Auric, who gets some utility A from destroying the currency, and Bond, who represents the existing participants in the Bitcoin economy and wants to preserve the value of his currency, initially valued at B. The game proceeds in two steps. First, Bond sets the mining reward C, which will cause ordinary miners to expend C on mining. Then Auric can decide to pay more than C to destroy the currency, or to do nothing. If Auric destroys the currency, he gets utility $A-C$ and Bond gets utility zero. If Auric does not destroy the currency, Auric gets utility zero and Bond gets $B-C$.

If $A > B$ then Auric will destroy the currency, because his desire to destroy it exceeds the amount Bond is willing to pay to save it. If $A < B$ then Bond will preserve the currency by setting the mining reward C equal to A, so that the miners do just enough work to make destruction unattractive to Auric. The key result is that survival of the currency requires the mining reward to be at least as large as Auric's utility from destruction. In effect, the Bitcoin economy must pay a "tax" of A to keep Bitcoin alive. The tax is paid in the form of resources expended on mining computations that otherwise have no value.

One obvious question is what happens when Bond is uncertain about Auric's utility function. We model this by drawing Auric's taste for destruction, A, from a known probability distribution with cumulative distribution F.

Suppose Bond bids x. Then the currency will survive iff $A < x$. Bond's expected utility is

$$\mathbb{E}(Bond's\ utility) = (B-x)F(x)$$

Bond will maximize his utility by choosing x such that:

$$\frac{d}{dx}\mathbb{E}(Bond's\ utility) = 0 = (B-x)F'(x) - F(x)$$

implying that

$$x = B - \frac{F(x)}{F'(x)}$$

if such an x exists.

As an example, suppose that A is distributed uniformly in the range from zero to A_{max}. Then if $A_{max} < \frac{B}{2}$, Bond will set the mining reward to A_{max} and preserve the currency with probability one. Otherwise, if $A_{max} > \frac{B}{2}$, Bond will set the mining reward to $\frac{B}{2}$ and save the currency with probability $\frac{B}{2A_{max}}$.

In general, if no suitable x exists, then the currency will die.

This corresponds to the "death spiral" scenario in which Auric can kill the currency by generating uncertainty about the possibility of an attack. As long as the threat is sufficiently credible, Bond cannot justify trying to save the currency. This in turn suggests that Auric can profit by bluffing: if Bond does not know Auric's utility and Auric can make strong claims about an imminent attack, Auric may be able to scare off rational players and start a death spiral without the need to mount a real (and expensive) Goldfinger attack.

6 Emerging Governance in Bitcoin

As discussed above, the success of Bitcoin relies on a consensus on the rules—on which transactions and blocks are considered valid and which are not. Contrary to claims that Bitcoin is ungovernable and relies on fixed rules laid down at its founding, the rules can be and have been changed by consensus. A governance structure for Bitcoin must inevitably emerge and is already emerging.

6.1 Bitcoin Governance in Practice

There are several examples of governance operating in the current Bitcoin system. One is a change in the minimum transaction size to discourage the creation of Bitcoin *dust*, or transaction outputs of very low value. Dust has been proliferating due to the behavior of the popular gambling game SatoshiDice and the use of the Bitcoin log as a global timestamping facility[11]. Originally, the minimum input to a transaction was defined to be 10^{-8} BTC, a unit known as a *satoshi*. The proposed update would change the minimum transaction input from 1 satoshi to 5430 satoshi (currently, about half of a U.S. cent). Smaller transaction inputs would then be ignored as invalid.

As another example, version 0.7 of the Bitcoin client reference implementation had a bug that caused it to reject as invalid a small fraction of log blocks that were in fact valid. This bug was fixed in version 0.8, and Bitcoin continued to operate with some clients running version 0.7 and some running 0.8. On March 11, 2013, a valid block was created in the Bitcoin log that triggered the 0.7 bug, so that 0.7 clients rejected the new block but 0.8 clients (correctly) accepted it as valid. The result was a fork in the log, with 0.7 clients creating a new branch which they saw as the only valid one, while 0.8 clients extended the branch containing the offending block. The situation is shown in Figure 3.

In order to prevent a long-term fork in the log, the community of Bitcoin miners determined that the 0.7 branch should be made the long-term consensus branch, even though it was the shorter of the two. This strategy was implemented by having some participants downgrade from version 0.8 to version 0.7, so that the 0.7 branch grew faster. After a time, the 0.7 branch became the longer one, allowing a return to the normal longest-branch-wins rule and allowing anyone to upgrade to 0.8 without re-creating the fork. Mining rewards on the 0.8 branch were forfeited [18]. The problem had been solved by governance: the mining community decided by consensus to make an exception to the rules.

6.2 Long-Term Need for Governance

The examples above illustrate why Bitcoin needs governance. Beyond this, the system will need governance to cope with longer-term structural challenges.

[11] A value can be timestamped by creating a low-value Bitcoin transaction that refers to the value, taking advantage of the fact that the Bitcoin log essentially timestamps every transaction.

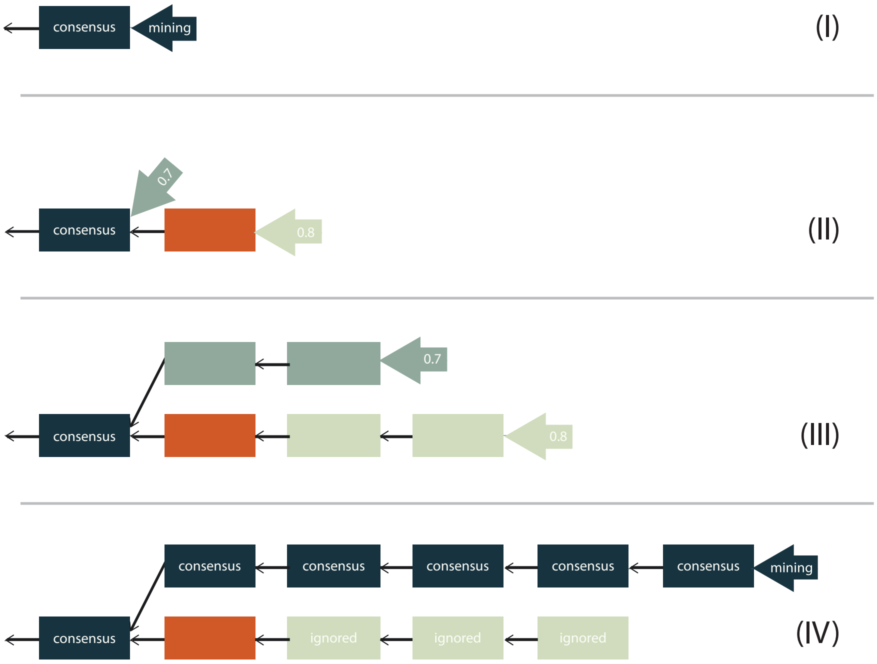


Figure 3: A schematic depiction of the March 11, 2013 blockchain fork. In (I), mining is proceeding as normal. In (II), a *bad block* is generated which causes a fork, shown in (III). Observe that the chain ignoring the bad block is shorter—if it were longer, clients accepting the bad block would simply shift to the longest branch. In (IV), enough miners have moved to the branch without the bad block to make it the longest branch; mining then proceeds as normal.

As an example, consider the need to maintain adequate mining incentives. If the mining incentive is inadequate, mining activity will shrink and Goldfinger attacks will be too easy. At present, the mining reward seems to be large enough, but under the current rules of Bitcoin the reward for mining will fall exponentially with time. Transaction fees, which are voluntary under the current rules, cannot make up the difference, as discussed in Section 4.2.

The only way to preserve the system's health will be to change the rules, most likely either by maintaining mining rewards at a level higher than originally envisioned, or making transaction fees mandatory. Different groups benefit from each solution (for example, raising the mining reward modifies the money supply, which is anathema to much of the Bitcoin community, but mandatory transaction fees can be seen as slowing adoption of the technology by merchants). The choice between relying on mining rewards versus relying on transaction fees amounts to an economic policy decision: whether to increase the money supply or to put a tax on transactions. In either case, a higher mining incentive would cause more resources to be expended on otherwise-useless mining activity.

The choice is likely to drive political disputes within the Bitcoin community. Some members believe strongly in maintaining a fixed money supply, and think that increasing mining rewards would debase or inflate the currency. On the other hand, a tax on transactions will harm those who rely on transactions while putting less burden on participants who buy and hold Bitcoins. A political choice such as this is difficult to make without some sort of governance structure, even if an informal one.

Other challenges to the system's health and viability may also emerge, perhaps due to issues of scaling or security. Some sort of governance will have to emerge in order to cope with these. Although it may be informal and not enshrined in any constitution or charter, the Bitcoin community will need to have a way to reach consensus decisions and act on them.

6.3 Emergence of Governance

Arguably, a governance structure is already emerging through the management of the Bitcoin reference implementation. The lead developers of this software are respected in the community and their opinions tend to carry weight. Because putting into practice any rule changes requires changing the reference software (and because the reference software is widely deployed), the lead developers have their hands on the levers of power, such as they exist. They seem to be the natural leaders of the community.

As with other open source projects, the power of the project's leadership is limited by the ability of anyone to *fork* the software by copying the current version and then evolving the copy separately. A fork will survive if it has enough support from the community, and it might even dominate the original version if there is a strong consensus for the new forked version. This possibility keeps the governance of the software mostly consistent with the desires of the community. Several forks of the Bitcoin software exist as less popular currencies with small variations in the rules.[12]

There is an entity called the Bitcoin Foundation [8], though it seems to be involved mostly in promotional activities rather than making decisions about the Bitcoin rules, so we conclude that the developers will have more influence in the long run.[13]

[12] For example, Litecoin [19] is a currency that disfavors professional mining by using cryptographic hashes that are not amenable to the use of specialized hardware, and Freicoin [14] is a currency that attempts to solve the problem of deflation in Bitcoin with demurrage. These currencies differ from rule forks described below in that they do not share any transaction history with Bitcoin - their transaction ledgers start with their own *genesis blocks*.

[13] Some of the lead developers are affiliated with the Bitcoin Foundation, so it might be difficult to separate the activities of the developers from those of the Foundation. Nonetheless, it appears that their influence comes mostly from their role as developers rather than their affiliation with the Foundation.

6.4 Forking the Rules, Forking the Currency

The *rules* of Bitcoin are subject to the same kind of open-source governance. In principle, anyone can fork the rules by announcing that at a certain time, they will consider the rules of Bitcoin to have changed in some way.

As with a software fork, a rules fork will only be sustainable if enough people adopt the new version. If this happens, the likely result would be a fork of the Bitcoin log, with one branch corresponding to each rule set. The log fork would occur the first time a log block is generated that is legal under one rule set but not the other. After that point, the followers of each rule set would stay on their respective branches. This is essentially what happened when the version 0.7 software bug described above in Section 6.1 caused two versions of the Bitcoin software to behave as if they had different rule sets.

If a rules fork occurs between rule sets A and B, and if both branches of the fork can sustain support in the community, then the currency will fork into two new currencies that we might call "Bitcoin A" and "Bitcoin B." Because the two currencies share a log up to the fork point, it will appear that the currency splits at the fork point, with an owner of one Bitcoin at the fork point receiving one Bitcoin A and one Bitcoin B. After the split, each currency would proceed separately along its own path.

As a result, the dynamics of Bitcoin's rules governance are similar to those of open-source software governance, with an emerging set of leaders who make decisions on behalf of the community and whose power is constrained by the possibility of a fork.

6.5 De Facto Governance

Finally, we note that in practice rules governance is entangled with governance of the Bitcoin reference software. Although many software changes are purely motivated by engineering factors unrelated to rules governance, the primary vehicle for actually changing the rules has been (and will likely continue to be) through changes to the reference software. Therefore, the lead developers of the open source reference software have become a de facto rules governance body for the Bitcoin economy.

7 Conclusion

Our analysis of incentives, stability, and governance in Bitcoin shows that Bitcoin is not the fixed, rule-driven, incentive-compatible system that some advocates claim. Although miners currently follow the original rules, this behavior is stable only by consensus and the rules could be changed at any time, either by a Goldfinger-style attacker or by a consensus governance process.

We also conclude that Bitcoin is more amenable to government regulation than advocates claim. The rules can be changed. They have been changed. And a semi-formal Bitcoin governance process is emerging. To the extent that Bitcoin's governance structure is subject to pressure from a regulator, or that a significant fraction of miners or users are subject to regulatory pressure, the regulator will be able to put pressure on the Bitcoin economy to change its rules.

Still, a regulator's power will be limited by participants' ability to fork the Bitcoin rules. Even if a regulator forces the developers to incorporate changes into the Bitcoin rules and reference software, the rest of the Bitcoin community will be able to fork the rules and carry on under the ruleset of its choice. Bitcoin is not immune to regulation, but it is not like traditional currencies either. Bitcoin is the first mainstream open-source currency.

Acknowledgments

We are grateful to Tim Bauman, Ari Feldman, Laura Felten, Mike Freedman, Alex Halderman, Tim Lee, Arvind Narayanan, Bryan Richter, Steve Schultze and Cameron Wilson for useful conversations and suggestions, and to the anonymous reviewers for their very helpful feedback. Kroll was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1148900.

References

[1] E. Androulaki, G. O. Karame, M. Roeschlin, T. Scherer, and S. Capkun. Evaluating User Privacy in Bitcoin. In *Proceedings of Financial Cryptography*, 2013.

[2] M. Babaioff, S. Dobzinski, S. Oren, and A. Zohar. On Bitcoin and Red Balloons. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 56–73. ACM, 2012.

[3] A. Back et al. Hashcash-a denial of service counter-measure. `http://www.hashcash.org/papers/hashcash.pdf`, 2002.

[4] S. Barber, X. Boyen, E. Shi, , and E. Uzun. Bitter to Better—How to Make Bitcoin a Better Currency. In *Proceedings of Financial Cryptography*, 2013.

[5] T. Bauman. Commerce and Reputation in Online Illegal Drug Markets. Princeton University Senior Thesis, 2013.

[6] J. Becker, D. Breuker, T. Heide, J. Holler, H. Rauer, and R. Böhme. Can We Afford Integrity by Proof-of-Work? Scenarios Inspired by the Bitcoin Currency. In *Workshop on the Economics of Information Security*, 2012.

[7] Bitcoin - P2P Digital Currency. `http://bitcoin.org`.

[8] Bitcoin Foundation. `https://bitcoinfoundation.org`.

[9] BTC Guild. `http://www.btcguild.com`.

[10] D. Chaum. Blind signatures for untraceable payments. In *Advances in Cryptology: Proceedings of Crypto*, volume 82, pages 199–203, 1982.

[11] W. Dai. b-money. http://www.weidai.com/bmoney.txt, 1998.

[12] C. Dwork and M. Naor. Pricing via processing or combatting junk mail. In *CRYPTO*, 1992.

[13] FBI Directorate of Intelligence Cyber Intelligence Section and Criminal Intelligence Section. Bitcoin Virtual Currency: Unique Features Present Distinct Challenges for Deterring Illicit Activity. http://www.wired.com/images_blogs/threatlevel/2012/05/Bitcoin-FBI.pdf, 2012.

[14] Freicoin - easy-to-use demurrage currency. http://freico.in/.

[15] G. Hamilton (Dir.). Goldfinger. United Artists, 1964.

[16] G. O. Karame, E. Androulaki, and S. Capkun. Double-spending fast payments in Bitcoin. In *Proceedings of the 2012 ACM conference on Computer and Communications Security*, pages 906–917. ACM, 2012.

[17] B. Laurie and R. Clayton. Proof-of-work proves not to work. In *Workshop on Economics and Information Security*, volume 2004, 2004.

[18] T. B. Lee. Major glitch in Bitcoin network sparks sell-off; price temporarily falls 23%. http://arstechnica.com/business/2013/03/major-glitch-in-bitcoin-network-sparks-sell-off-price-temporarily-falls-23/.

[19] Litecoin - Open source P2P digital currency. http://litecoin.org/.

[20] I. Miers, C. Garman, M. Green, and A. D. Rubin. Zerocoin: Anonymous Distributed E-Cash from Bitcoin. *IEEE Symposium on Security and Privacy*, 2013.

[21] Mt. Gox - Bitcoin Exchange. http://mtgox.com.

[22] S. Nakamoto. Bitcoin: A Peer-to-Peer Electronic Cash System. http://bitcoin.org/bitcoin.pdf, 2009.

[23] J. Ochs and A. E. Roth. An experimental study of sequential bargaining. *The American Economic Review*, pages 355–384, 1989.

[24] organofcorti. Neighborhood Pool Watch. http://organofcorti.blogspot.com/.

[25] N. Popper and P. Lattman. Never mind Facebook; Winklevoss twins rule in digital money. http://dealbook.nytimes.com/2013/04/11/as-big-investors-emerge-bitcoin-gets-ready-for-its-close-up/.

[26] D. Ron and A. Shamir. Quantitative Analysis of the Full Bitcoin Transaction Graph. In *Proceedings of Financial Cryptography*, 2013.

[27] T. C. Schelling. *The Strategy of Conflict.* Harvard University Press, 1960.

[28] C. E. Schumer and J. Manchin. Manchin urges federal law enforcement to shut down online black market for illegal drugs. `http://manchin.senate.gov/public/index.cfm/press-releases?ID=284ae54a-acf1-4258-be1c-7acee1f7e8b3`, 2011.

[29] B. Skyrms. *The stag hunt and the evolution of social structure.* Cambridge University Press, 2003.

[30] P. Wuille. Bitcoin network graphs. `http://bitcoin.sipa.be/`.