Beyond Bitcoin

Economics of Digital Currencies and Blockchain Technologies

Second Edition

Hanna Halaburda · Miklos Sarvary
Guillaume Haeringer

palgrave
macmillan

Beyond Bitcoin

"Digital currencies and Blockchain technologies have the potential to revolutionize finance and commerce, yet they're widely misunderstood and overhyped. Halaburda, Sarvary and Haeringer provide a clear-eyed guide to the potential and pitfalls of these technologies in their well-written and carefully-researched book."

—Erik Brynjolfsson, *the Jerry Yang and Akiko Yamazaki Professor at Stanford, USA and Director of the Digital Economy Lab; co-author of the book* Second Machine Age

"While blockchain currencies such as Bitcoin or Ethereum hold the promise of reshaping economic activity and transforming our lives through increased control, security, and a reduction in the costs of carrying transactions, for many, the world of digital currencies and blockchain remains dark, puzzling, and mystifying. *Beyond Bitcoin* lays out in a clear and concise manner the economics of digital currencies and blockchain technologies, and should be read by anyone seeking to benefit from the opportunities that this quickly evolving phenomenon presents."

—Ramon Casadesus-Masanell, *Professor, Harvard Business School, USA; co-editor,* Journal of Economics & Management Strategy

"*Beyond Bitcoin* is an instant classic. It demystifies digital currencies and blockchain, and should be required reading for students, managers, investors, and anyone interested in the economics of digital money and these new technologies."

—Annabelle Gawer, *Professor and Director of Center for Digital Economy, University of Surrey, UK; co-author of the book* Platform Leadership

"This book is a crystal-clear introduction to the power and potential of cryptocurrency – a must-read for anyone hoping to understand how and why it is changing the world. Halaburda, Sarvary, and Haeringer masterfully situate crypto within the history and economic foundations of money, and explain what makes it so different from previous attempts at digital currency. They then teach how different cryptocurrency platforms such as Bitcoin and Ethereum work. And finally, drawing heavily on their expertise in platform economics and market design, the authors give a glimpse into which crypto applications are likely to be here for the long term – and where the biggest opportunities are."

—Scott Duke Kominers, *MBA Class of 1960 Associate Professor, Harvard Business School; Faculty Affiliate, Department of Economics, Harvard University, USA; columnist,* Bloomberg Opinion

Hanna Halaburda · Miklos Sarvary ·
Guillaume Haeringer

Beyond Bitcoin

Economics of Digital Currencies and Blockchain Technologies

Second Edition

palgrave
macmillan

Hanna Halaburda
Stern School of Business
New York University
New York, NY, USA

Miklos Sarvary
Columbia Business School
Columbia University
New York, NY, USA

Guillaume Haeringer
Zicklin School of Business
Baruch College
New York, NY, USA

ISBN 978-3-030-88930-2 ISBN 978-3-030-88931-9 (eBook)
https://doi.org/10.1007/978-3-030-88931-9

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Gabriela and Lena,
who are changing and growing as quickly as blockchain technologies.

—Hanna Halaburda

To my wife.

—Miklos Sarvary

To Anna, Circe, Hermes, and Nora,
por todo.

—Guillaume Haeringer

Preface

Beyond Bitcoin has emerged from endless discussions between Hanna and Miklos that go back to 2014. Back then, "cryptocurrency" really meant "Bitcoin" and it appeared that it might remain only a marginal chapter in the digitization of finance. Instead, the lead players were large platforms that started experimenting with centralized digital currencies to further leverage their huge audiences. The book's title reflects our prediction at the time that centralized currencies by private platforms are likely to have a dominant role in the future of finance. In subsequent years, cryptocurrencies have come back with a vengeance, with variety driven by unprecedented innovation but also with clearer economic significance as an asset class, as a payment system, or as a serious environmental concern. This major change has sparked renewed interest in the book, leading our Editor to ask us to write a second edition.

It was immediately clear that the second edition of the book needed a major upgrade because a variety of innovations have completely changed the crypto landscape. Important new developments, such as ICOs and smart contracts had to be addressed. Moreover, academic research on cryptocurrencies has exploded in the last five years. We needed new expertise and we were fortunate that Guillaume joined our team. He brought a new perspective on the industry and a deep knowledge of relevant academic research. We were happy to see that our discussions haven't lost any of their intensity, even though most of them were conducted over Zoom during the Covid-19 pandemic.

The world of cryptocurrencies is far from a stable, settled state. On the one hand, we all realize the importance and the future potential of cryptocurrencies in complementing or even renewing the world's financial plumbing. On the other hand, one cannot ignore the systemic risk that cryptocurrencies may represent for the world economy. This risk has not escaped the attention of governments all over the world. Regulators everywhere are trying to implement a framework that balance the need for financial stability with the possibility of further innovations. It is in this extraordinary context that, we believe, the second edition can help interested readers better understand the economics of digital currencies.

New York, USA — Hanna Halaburda
New York, USA — Miklos Sarvary
New York, USA — Guillaume Haeringer

Acknowledgements

Hanna Halaburda
No ideas are born in isolation, and there are many people who helped me refine mine in the context of digital currencies and blockchains. I am thankful to Neil Gandal for embarking with me on an early empirical investigation of cryptocurrencies; to Joshua Gans for embarking with me on the first investigation of the design of digital currencies and their role in platforms strategy; to Kamil Niemira for his ongoing guidance and encouragement, and for helping me understand the audience. The first edition of the book would not have been possible without the support of the many wonderful people at the Bank of Canada, including Jason Allen, Kim Huynh, and especially Ben Fung. The continuous discussions and academic investigations with Yannis Bakos and Natalia Levina turned out invaluable in shaping ideas for the second edition. My special thanks go to Arthur Zwiegincew, who helped me understand the connection between academic research and real-life applications of technology. Last but not least, a warmest *thank you* to my family for their patience and support throughout this and all other adventures.

Miklos Sarvary
I am indebted to INSEAD for supporting my research on metaverses that originally sparked my interest in digital currencies. Discussions with colleagues both at INSEAD and later at Columbia Business School were invaluable to clarify my thinking on the subject. Finally, I have to thank my wife who patiently supported my work during all these years.

Guillaume Haeringer
I am immensely indebted to Hanna Halaburda for initiating my interest in cryptocurrencies and blockchain technologies. Her ability to always favor questioning over wonder was an undeniable tool while writing this second edition. I am grateful to Yannis Bakos, Ayan Bhattacharya, Agostino Capponi, Neil Gandal, Joshua, Aija Leiponen, Alex Teytelbom, and Llewelyn Thomas for their conversations, insights, and comments. Of course, all this would not have been possible without the encouragement from my family.

Contents

List of Tables

CHAPTER 1

Introduction

For more than two decades, the Internet combined with the smartphone revolution has created a permanently connected world transcending national borders, time differences, and geographic distance. In this way, the Internet has become the backbone for most of our activities, communication, and entertainment—of course—but also for much of our economic activities, e.g., work, consumption, commerce, etc. By the 2010s, it has actually become difficult (if not impossible) to live without using the Internet one way or another.

The technological progress also affected how we store our money and how we pay for the goods and services we need. In many countries people use less and less cash, making all their transactions electronically through credit cards, bank transfers, or using payment services like Paypal or Venmo. Those changes have affected our perception of what money is—or can be—and for the past two decades we are experimenting with types of money that have not been seen before in human history. These digital currencies only live in the virtual world of the Internet, are governed by often unfamiliar rules, and require us to adopt new habits if we want to use them. Some of the digital currencies come from issuers we are intimately familiar with, for example, social networks such as Facebook or commerce platforms such as Amazon. More often, they belong to the curious group of cryptocurrencies: digital currencies that have no person

H. Halaburda et al., *Beyond Bitcoin*,
https://doi.org/10.1007/978-3-030-88931-9_1

or institution managing issuance, no authority regulating them, and that operate throughout a decentralized peer-to-peer network.

Cryptocurrencies are the driving force behind people's interest in digital currencies. They certainly deserve the attention because of the technical innovation they represent. Bitcoin involves a sophisticated algorithm that managed to solve a long-standing puzzle, that of assuring consensus in a decentralized, permissionless network. While few people have heard about the problem, most of us have heard about the first operational solution, Bitcoin which was introduced in 2008 by a mysterious character called Satoshi Nakamoto, whose real identity is unknown. This solution offers the tantalizing possibility of payment systems and currencies, that operate in a distributed network, with no issuer or institution that controls or manages it, and with enough security to withstand most malicious attempts to infiltrate it. As it is widely discussed, this innovation has the potential to meaningfully change the economy, from the way cross-border remittances are sent, to making micro-payments economically sustainable, to offering a way of transacting online that protects privacy better than any other method.

The timing of Bitcoin's innovation could not have been better for it to attract popular attention. Around the same time, the world experienced the largest global financial crisis in modern history. The crisis led some people to question the management of national currencies and the institutions involved in it, in particular the financial sector and the government. Many people then felt that the time had come for the creation of a payment system that is safe, practical for global economic interactions, and, importantly, independent of existing large financial institutions and national governments. This independence would also alleviate increasing worries about government surveillance. Some people felt the need for a payment system that is, simply, out of governments' sight. Indeed—for better or worse—the last twenty years have seen a general increase of government control over citizens pretty much everywhere in the world.[1] This is reflected in more stringent monitoring of citizens' activities (including economic activities) and, also, in more stringent regulation, limiting these activities.

Beside the need for independence, there was also an economic rationale for embracing such an alternative payment. Bank transfers, especially

[1] See *The Economist*'s front page in August 3, 2013, entitled: "Liberty's lost decade."

international ones, are expensive and inflexible, imposing unreasonable costs on individuals and companies. Again, financial institutions and regulation have been largely blamed for these inefficiencies.

Since the Internet became popularly accessible, some people naturally saw—and see even today—the ubiquity of the Internet and the quasi-independence of its infrastructure from governments as the ideal technological combination for the introduction of a new, decentralized infrastructure. Cryptocurrencies and blockchain-based smart contracts may offer the missing piece of technology to achieve this goal. Bitcoin's design was so novel that initially few enthusiasts believed that it would actually work in practice. The fact that Bitcoin has functioned for several years without failure has been considered by many a proof of the validity of its novel design, and defence against sceptics. This created a growing enthusiasm about the principles of Bitcoin's design.

As a result of this broad enthusiasm, the last decade has seen a continuous flow of innovation in the domain of digital currencies and blockchain technologies. First, the crypto space had experienced the introduction of a large number of cryptocurrencies with similar designs. Subsequent developments went in more complex directions with blockchains focusing on running single programs, called smart contracts, then running decentralized applications built on these smart contracts, and possibly whole decentralized organizations. The first blockchain allowing for such complex designs was Ethereum, and decentralized structures that have been since set up on Ethereum include trading systems, like Uniswap, or investment funds, like The Dao. As with any new technology, there are successes and failures. Nonetheless, the developments in this space continue.

The developments were not limited to the pursuit of full decentralization and independence from traditional infrastructure. Blockchain technologies have been as well adapted to the benefit of large established companies like IBM, Maersk, or Walmart. These blockchains significantly differ from Nakamoto's design for Bitcoin. While Bitcoin is completely open, the enterprise blockchains are closed and actively managed structures that allow independent companies coordinate and unify information and processes automatically. Despite the differences, the development of enterprise blockchains also owes to the popularity of Bitcoin.

The universe of digital currencies also goes beyond that of cryptocurrencies, and into the world of more traditional industries. Indeed, a whole new family of digital currencies has emerged in parallel to Bitcoin and the

cryptocurrencies. The rise of these currencies is also closely linked to the widespread adoption of the Internet. But in contrast to Bitcoin's libertarian motivation, the creation of platform-based digital currencies has been motivated by the needs of a new breed of large Internet businesses: Amazon, Facebook, Tencent, etc.

Permanent and ubiquitous connectivity provided by the Internet has given rise to new business models that take advantage of a very large number of people interacting in sophisticated ways. Social networks, e-commerce platforms, online game platforms, or virtual worlds are so-called "transaction platforms," that create value by facilitating exchange between their members, who often represent different groups of consumers (buyers, sellers, advertisers, or developers). The nature of the exchange, whether it is social/commercial or whether it is for entertainment or concerns a particular professional/business purpose often defines the business model of the platform, including its value proposition and the way the platform earns its revenue. While these value propositions and revenue models substantially vary across transaction platforms, quite naturally, most of them provide the possibility of economic exchange between their members and between these members and the platform itself. This raises the question of the necessity of a medium of exchange, essentially an efficient payment system, maybe tailored to the special needs of the platform. Many platform businesses have considered introducing a special currency to provide one. As opposed to cryptocurrencies—where the goal has always been the creation of a decentralized currency—platform-based currencies are, by definition, centralized currencies where the platforms control (to the extent possible) the "rules" governing the use of their currencies.

Platform-based currencies periodically captured the public imagination just as Bitcoin did, no doubt partly because of these platforms' sheer size and global nature. For example, when Facebook was moving forward with their Facebook Credits in 2011, commentators saw them as a threat to traditional currencies. "Could a gigantic non-sovereign, like Facebook someday launch a real currency to compete with the dollar, euro, yen and the like?" wrote Matthew Yglesias (2012). Similarly, renown payments economist, David Evans (2012) wrote: "Social game companies could pay developers around the world in Facebook Credits and small businesspeople could accept Facebook Credits because they could use them to buy other things that they need or reward customers with them. In some countries (especially those with national debts that are greater than

their GDPs) Facebook Credits could become a safer currency than the national currency." Similar concerns were expressed when Amazon introduced Amazon Coins in 2013. The *Wall Street Journal* wrote: "But in the long term what [central banks] should perhaps be most worried about is losing their monopoly on issuing money. A new breed of virtual currencies are starting to emerge – and some of the giants of the web industry such as Amazon.com Inc. are edging into the market."[2]

As we know now, most of these concerns were exaggerated. In their time, these concerns gained ground because of insufficient understanding of how the technology works and what were the objectives behind introducing it. The core issues guiding the introduction of the platform-based currencies are very different from those of cryptocurrencies. While in the latter case, the goal was to create a fully functional currency to *replace* traditional fiat currencies, platform-based currencies try to purposefully design their payment systems with specific objectives in mind. This usually, but not always, boils down to restricting some of the functionalities of their currencies.

The goal of this book is to explore the young and dynamic universe of digital currencies and blockchain technologies to understand their origins and meaning for our economies. We approach these phenomena from the viewpoint of economists, analyzing the needs they fulfill for customers and merchants, the incentives they create for their users, and the way they compete. Whenever possible, we will do that in a way that abstracts away from technical details of how digital currencies and blockchains work, making this book suitable for people with little experience or education in computer science, cryptography, etc. Sometimes we won't be able to avoid talking about technical aspects—for example, we could scarcely avoid discussing the ingenious algorithm that underlies cryptocurrencies such as Bitcoin—but we will attempt to do so in a way that is as approachable as possible. Rather than creating a technical manual, we intend to describe the economic forces governing the evolution of digital currencies and blockchain technologies. The objective is to understand why certain designs seem to succeed over others and what design features (or restrictions) make sense in given economic or business contexts.

To this end, we will start at the very beginning. In Chapter 2 we will describe how human societies invented money, how money facilitated

[2] See WSJ Market Watch, "With Amazon minting currency, Fed at risk," www.marketwatch.com/story/could-amazon-run-central-banks-out-of-business-2013-02-13.

transactions, and how weaknesses in the design of money led to innovation and improvements in the way we pay for things. It might seem surprising to start a book on digital currencies with a chapter on history, or even pre-history, of money. However, this historical overview allows us to identify some of the core economic forces that drive the use of "money," highlight the specific needs that money serves, and illustrate the key attributes that money should have. These needs and attributes are remarkably universal, and they are as important now as they were centuries ago. Their analysis will lay the ground for our subsequent discussion of digital currencies and give us a framework in which to analyze them.

Such a framework is critically important. Without it, what exactly is going on in the digital currency universe can be difficult to understand. Much of the narrative surrounding digital currencies is a bit sensational, undoubtedly influenced by the tumultuous events surrounding the introduction of digital currencies, the spectacular developments of Bitcoin (its rise to immense popularity, but also the less optimistic episodes of the Silk Road shutdown or the closing of the Mt. Gox exchange). Starting with an economic framework will help us see through the confusion to better understand the phenomenon of digital currencies and its potential to change our economy.

In Chapter 3, we will use this framework of money attributes to explore the universe of platform-based digital currencies that are centrally managed by the businesses that have introduced them. We will discuss the economic forces that made it attractive for Amazon to issue the Amazon coin, or for Facebook to issue Facebook Credit—and why the companies decided to shut them down after a while. Here, we will also attempt to address the challenge of figuring out what drives the platform's choice of particular design features for its private currency.

One of the main messages in this part of the book is that platform-based digital currencies could scarcely function as money in the broad sense of the word—not because they are inherently flawed, but because platforms issuing them go to great pains to disable the main functions that are necessary for a widely adopted currency. We will see that this should not be surprising: such restrictions fit well with the platforms' business model and make their currencies more useful in generating a higher profit for the platform. At the same time, platforms can extend their scope and turn to new business models implementing cryptocurrencies beyond their main platform functionality.

A widespread adoption, and perhaps even crowding out traditional currencies, is something often discussed in the context of decentralized digital currencies, or cryptocurrencies such as Bitcoin. We discuss these innovations in Chapter 4. We also look into the ecosystem that cryptocurrencies exist in, focusing on its more economically meaningful parts. And while we acknowledge the ingenuity of Bitcoin's design, we also elaborate on unintended consequences of the incentives induced by the Bitcoin protocol.

Chapter 5 is devoted to the evolution of cryptocurrencies in the context of economic forces. While Bitcoin and its immediate competitors had a unique objective to offer a peer-to-peer electronic currency, the innovation market quickly shifted towards differentiation in terms of functionalities cryptocurrencies could provide. The demand for the cryptocurrency was not driven by the quality of its design, but rather how and where it could be used. The development of Ethereum considerably changed the game by bringing smart contracts and crypto-tokens to the realm of cryptocurrencies, and thus making issuing new "coins" easier and cheaper.

Ethereum introduced a new paradigm, where the blockchain is viewed as a platform whose purpose is not only to manage a cryptocurrency, but that can also host other coins (tokens) or applications through smart contracts. Chapter 6 is devoted first to a careful presentation of Ethereum, emphasizing the motivation that led to its creation and how it differs from the Bitcoin design. We then discuss how smart contracts, decentralized applications, and, in particular, crypto-tokens create new competitive opportunities, and even create new markets.

Businesses and other institutions often need to manage distributed data, that is, data that is shared and maintained by different parties. And thus, they may find blockchain design attractive. We discuss in Chapter 7 the advantages and downsides of blockchain for enterprise use and how the use of permission rights can affect blockchains' design constraints. In particular, we explain how permissioned blockchains, if used in networks with a certain level of trust, can be a tool to address a wide range of problems.

Such discussions often turn into speculation about the future, a temptation we have not managed to resist. At the same time, we clearly recognize that it is too early to paint an exact picture given the broad-scale experimentation still under way. More importantly, such forecasts are particularly difficult in the light of the uncertainty about how governments will respond to the activities related to the digital currencies and application of blockchain technologies. In this respect, our book is not a

policy piece about central banking or currency regulation. Rather, it is an analysis of the economic forces that drive the emergence and efficient use of such new technologies.

CHAPTER 2

Means of Exchange: Ever-Present Competition

Digital currencies are only a recent innovation and their widespread use for payments is still a thing of the future. It is fit, however, to begin our investigation with looking into the past. In fact, we'll start with a time well before the digital era, in fact, before the development of money. We do not intend here to provide a comprehensive overview of the history of money.[1] Rather, in our discussion we will focus on the characteristics (attributes) of different currencies, and various economic needs that money serves. In historical perspective we can see competitive forces that make some means of exchange more successful than others in satisfying those needs. We will later see that digital currencies can be successful only if they satisfy those needs as well as or better than the traditional currencies we already have in use.

We'll also overview the various objects and technologies that have served as money or, more broadly, as various means of exchange. We'll see examples of coexistence of various currencies, episodes that suggest a tantalizing possibility that, in the future, digital currencies may coexist alongside other digital currencies, but also side by side with traditional money. We can also identify when such possibility exists.

[1] For a comprehensive history, see e.g. Martin (2014), Fergusson (2008), Weatherford (1997).

H. Halaburda et al., *Beyond Bitcoin*,
https://doi.org/10.1007/978-3-030-88931-9_2

Finally, we'll talk about competition between different currencies. Again, this discussion will yield useful insights later. For example, digital currencies are being introduced alongside, and necessarily must compete with, traditional money. Eventually, if digital currencies win more widespread adoption, we may need to turn these arguments around and use them to discuss whether traditional money can survive in the long term in the presence of digital currencies.

With this roadmap in mind, let's move on and start with a brief history of how we trade.

2.1 The Medium of Exchange—Historical Overview

If you asked your friends why modern economies need money, they would most likely answer "to buy things." This answer would be as simple as it is deceptive. It is certainly true that we need money to facilitate trade, but there was a time in human history when transactions occurred without any money. Most of us have heard about barter, exchange of a product or service directly for another product, without the use of money. But the first economic transactions likely predate even that development. In essence, they were based on trust.

There was no need for money in the pre-agrarian hunter–gatherer groups.[2] The members of the group were all responsible for a communal provision of goods. The group kept track of each member's contribution and imposed penalties to minimize potential free-riding.

The collective memory of the group served as a ledger, or perhaps a prehistoric bank account. Members who contributed to the well-being of the group could count on being reciprocated in the future. The side benefit of this simple but ingenious arrangement was credit. A member of a group could potentially count on receiving goods and services even if he or she hadn't yet earned enough "brownie points" to justify them. As long as the group remembered about the transaction, they could expect the member to repay it with good deeds in the future. If they didn't, then the group could presumably discipline them by not allowing them to participate in the system going forward.

[2] See Harari (2014).

Of course, counting on collective memory only works if the group is of a relatively small size. Over time, groups grew larger; for example, people started settling in early cities. Eventually, people were unable to keep track of individual contributions. Moreover, as different groups started trading with each other, it became necessary to trade with less-known people whose prior contributions were unknown and who could hardly be disciplined to repay for a product in the future.

Without the help of collective memory and group-imposed discipline, transactions became risky: you could no longer be certain that people you trade with would repay you in the future. Nonetheless, there are gains to trade, and when people see sufficiently large gains, they will find a way to make trade happen. So, this did not stop transactions completely but forced them to be based on immediate exchange of goods for goods: barter.

Barter would work very well—as long as you find a seller offering something you want, and at the same time you have something the seller wants in return. In practice, this "double coincidence of wants" may happen fairly infrequently.[3] This is an important problem that limits trade. If you want to obtain a particular good, it may already be difficult for you to find somebody who has that good to offer; it will be even rarer that you might have something that person wants in return. You may need to rely on longer chains of buyer and seller (to get something Alice wants, I need to trade with Bob first), but of course it is even more difficult to find three or more people with suitably aligned holdings and wants, and get them to come to the same place at the same time. Affecting all trades at the same time is safer. With more parties it may also be more risky in the sense that the first person to hand over their good is the last one to receive the traded good. There is risk that something will go wrong along the way, and the first person in the chain may lose their original good and not receive much in return.

Barter had one more drawback: the timing coincidence. For example, many products are seasonal and may be difficult to store for longer periods of time. In the fall, you may have some berries that you'd be happy to exchange for meat when the winter comes—but since winter is still a few months away, you won't be able to exchange the goods in a pure barter transaction. So, in a barter transaction, the two sides not only

[3] See Jevons, W.S. (1875), Kiyotaki and Wright (1989).

need to want their respective goods, but they need to want them and have them available at the same time. Because of such frictions, many potential trades may not occur, leaving the parties that would have benefited from trade worse off.

As societies grew larger, and as new trade opportunities between various groups arose, these frictions and the foregone benefits of trade increased. Emergence of money in such a situation is not inevitable, but the potential benefits may have eventually become so large that they could no longer be ignored. This illustrates the main function of money, intuitively obvious to most people: money is there to facilitate trade, to overcome the double coincidence and timing problems and allow us to obtain the goods and services that we need.

Those early societies which coordinated on using tokens or intermediate goods had more opportunities to trade. In fact, there are no documented large societies using pure barter. Using intermediate goods for trading, however, dates back to at least 3,000BC. The earliest kinds of such intermediate goods were related to foodstuffs such as barley.

Using popular foodstuffs helped alleviating the first problem, the coincidence of wants. Everybody in a society consumed similar foodstuffs, making them a product that was attractive to all society members ("we all could use more barley"). Of course, the innovation was that people started accepting barley not only for their own consumption, but also in expectation of using the barley for other future transactions. It is likely that this innovation was not decreed by a ruler ("we will all use barley as money") but rather occurred organically. In either case, money, in the sense we typically mean it nowadays, was born. Except that this earliest money also served another useful role: it was food.

The foodstuffs used to facilitate trade differed across societies. Barley, likely the first historical example, was used in ancient Mesopotamia. Salt has been used in China in thirteenth century and in Ethiopia from sixteenth until twentieth century, whereas the Aztec Empire adopted cacao beans. All these examples share some important traits. First, they were relatively uniform and easy to divide. One can make smaller or larger units, by weight or volume (barely), by breaking smaller and larger pieces (salt), or collecting smaller or larger amounts (cacao beans). If you measure barley using a standard cup, that cup will hold a similar amount of the foodstuff currency this year and the next, at home or in a neighboring village.

Only relatively durable foodstuffs were adopted as money. In this, they clearly dominated other articles of food such as perishable fruit, fish, or milk. Nonetheless, they could not be stored indefinitely, and sometimes lasted only for one or at most a few seasons. Foodstuffs deteriorate quickly for a range of reasons. The foodstuff money could perish when exposed to the elements or, more prosaically, could be eaten by animals.

The money kept evolving and, around 1,200BC, an innovation appeared: money based on tokens that were not related to food. Perhaps the most well-known of such tokens were cowry shells, in widespread use in Africa for hundreds of years. The range of such money was, however, much greater. To give just two more colorful examples, until the twentieth century dog teeth served as money in the Admiralty Islands and, on Fiji, whale teeth played the same role until the nineteenth century.[4]

This token-based money had clear advantages relative to foodstuffs. The tokens would keep for much longer than one season. They were also easier to store or transport over longer distances. An important feature of token-based currencies was that the tokens represented value in a more abstract, symbolic way than barley or cacao did, as they had less intrinsic value than food. Usually they had cultural meaning and were also used for decoration. Interestingly, it is not clear whether they developed into currency *because* they had cultural meaning, or they gained the meaning because they could be used in exchanges, and therefore represented more value.

Along these advantages, there were a few distinct drawbacks. While foodstuffs that were used as currencies were relatively uniform, the tokens used as money varied greatly in shapes, sizes, and colors. These differences, naturally occurring in shells, teeth, an so on, made it more difficult for people using the currency to agree on which "prices" they represented. For example, a fish might be worth three dog teeth, but perhaps the seller would demand four teeth instead if the teeth were particularly small. In some cases, such differences between tokens were used to the advantage. For example, on the Yap Island, blue-lipped cowrie shells, a rarer type than the more popular yellow-lipped kind, served as a "higher denomination" currency.

A particular type of token money were metal pieces. The first use of metal as currency that we know about occurred in ancient Mesopotamia,

[4] See Einzig (1966).

2500 BC.[5] Metal proved to be even more durable than shells or teeth. It was also easily divisible into smaller units, and these units could be directly compared to each other based on their weight. This represented an improvement over naturally occurring shells or teeth.

Nonetheless, metals had not completely solved the problem of non-uniform units. While it was easy to weigh pieces of metal, there were several types of metals in common usage: copper, silver, and, of course, gold. Moreover, even one type of a metal may have differed in purity. These differences led to difficulties and additional risks in conducting transactions, particularly when metal money was used by people without specialized knowledge about it. Risk around the value of received payment would make some sellers wary, and they may avoid some trades that could otherwise be beneficial.

The problem of non-uniform units was the likely driver of the next innovation: metal-based coins. These uniform pieces of metal, with a stamp indirectly certifying weight and purity, represented uniform units. Two coins with the same stamp were considered equivalent; different stamps were readily recognized as agreed-upon indicators of the weight of the coin or type of metal. This made transactions (exchanges of metal for goods) much easier. One did not have to have scales handy or know how to use them; or have expertise to judge metal purity. One could rely on the stamp as the indicator of value.[6] This, of course, worked well when people trusted the stamp. Typically, mints would be directly or indirectly controlled by the sovereign. The benefit of the coin then depended on people's trust in the authority and integrity of the ruler. When people did not trust the stamp, they reverted to older methods of weighing and checking the purity of the metal.

The first coins of this type were introduced in the kingdom of Lydia in the seventh century BC. They were minted from electrum, a naturally occurring mixture of gold and silver, but the silver and gold coins soon

[5] See e.g. Weatherford (1997).

[6] That, of course, did not stop people from trying to get an unfair edge in a transaction. The most obvious example of dishonestly manipulating coins is debasement. It is not certain when this procedure started, although some sources point to the reign of the Roman emperor Nero (see Comparette, 1914). In the face of such manipulation, people weighted the coins again. It was only king's illusion that this would solve anything, because people used the underlying value of the metal to assess the value of the coin. But even then, at a given time, the value of a particular coin was well-known, and trade was easier than with random pieces of metal.

followed. An interesting innovation was that Lydian coins were relatively small, making it easier to store and transport the currency. Each coin was worth a few days of laborer's work, or a small part of a harvest. This opened up what could be called retail market to more trading opportunities.

The Lydian invention turned out to be more attractive than the earlier types of money. The invention quickly spread throughout Mediterranean, and metal coins of different value and sizes became the main tool of the trade in the western world until the Renaissance. The basic model remained unchanged until now. Coins are still metal discs with a stamp certifying the value of the piece (traditionally, also the purity and the weight of the metal). Of course, some improvements were introduced over time. For example, to prevent debasement, coin edges were stamped or rimmed, making it easier for users to identify whether pieces of metal were cut from a coin, changing its weight and its value.

The next significant innovation in money was paper money. Historically, it was first introduced in China in the eighth century. It is possible that the idea of paper money was brought to Europe by Marco Polo. In Europe, paper money became popular during the Renaissance, when Italian bankers introduced bills of credit. Both in China and in Europe, paper substituted for metal because it was cheaper, easier, and safer to transport. A person carrying paper money was less conspicuous than a carriage with valuable metal. Therefore, carriers of paper money were less likely to be attacked on the roads. Both because of lower risk of attack and because only the person carrying the paper needed guarding and not the carriage with metal pieces, one needed to hire fewer guards to travel safely than when transporting the same value of metal.

For several centuries paper money represented a claim on metal money. It was done through different types of promissory notes. Receipts for deposits are the simplest one. When a person deposited gold with a Renaissance goldsmith-banker, he (usually a he at that time) would get a receipt. With this receipt the gold could be withdrawn from the goldsmith. Originally the receipts were personal, but later they became payable to the bearer. That allowed for transferability, and thus the receipts could be used in transactions in lieu of the gold itself.

Later, when banks started issuing the bank notes, holding a dollar note from the Bank of Augusta, Georgia meant that the Bank of Augusta

would at any time redeem that note for specie (gold or silver coins). It was true, in principle, until gold standard was abandoned in 1970s.[7]

After the gold standard was dropped, paper money was no longer a claim on metal or any other good. It became fiat money, money "on the say-so." Countries made them legal tender, in the sense that they were accepted as payment of taxes and debts. Merchants needed to accept it, unless they explicitly state that they do not. But most importantly, paper money is accepted because the sellers know they can spend it as money. It has no intrinsic value, unless you count recycling value of the paper. Their value is purely symbolic.

It is true not only about paper money. Even though metal seems to have more intrinsic value than paper does, modern coins' value derives from the number stamped on them. The coins are no longer minted of gold and silver, but of less-valued metals like copper and nickel. Most coins have the symbolic value larger than the value of metal in them. But in some cases, it costs more to make them than their face value. For this reason Canadian Mint stopped issuing one cent coins in 2013.

While today we accept paper money as one of the most common forms of currency, it was not the case historically. There were often problems in introducing paper money, for example, because the populace did not consider it as trustworthy as metal coins and possibly feared overissuance. In some regions, notably China, paper money representing metal money was introduced successfully because it was imposed and guaranteed by the state. The state in fact resorted to executing people who refused to comply, and to confiscating other potential means of payment, like metal or gems.[8] In Europe, paper money had more difficulties in becoming generally adopted. European states did not impose as strong of an enforcement, and neither did they guarantee the paper money's value. There were several cases of governments overissuing paper money that later was not redeemed at the promised value.[9] This created mistrust of paper money and hindered its widespread adoption.

[7] In practice, however, in the United States a set of rules prevented the actual redemption after 1933, as owning gold was illegal for Americans between 1933 and 1971.

[8] See Weatherford (1997).

[9] For example, French Banque Royale in 1716–1720, issuance of Continentals during American Revolution, or Confederate dollars during the Civil War in US.

The final development, which brings us to the current times, is electronic money it.[10] Most often, when people think of electronic money, they think of credit cards. Credit cards, however, did not start as electronic. They did not even start as plastic. They started as cardboard cards back in the 1950s. Credit card systems are based on a ledger. Transactions are recorded and reported to an institution holding the ledger and the accounts. The institution, usually a bank, checks whether the funds about to be spent are available, bundles transactions for billing the account holder, and usually also offers credit services, allowing to defer payment to the future. The credit card gives information about the account and the system where the account is held. Introduction of digital technologies allowed for electronic report of the transactions, sped up authorization, and decreased fraud.

The widespread adoption of the more modern forms of money—metal coins, paper money, increasingly electronic money—is driven by their advantages over earlier forms. However, even nowadays, there are still situations when the earlier types of money reappear. For example, the shortage of the usual currency in the prisoner of war camps led to the use of cigarettes as currency. The same mechanism, and the same token, has been adopted as the currency in the informal economies in prisons. Interestingly, when smoking was banned in some prisons, cigarettes disappeared, but money did not: prisoners started using cans of mackerel as currency.[11] Table 2.1 summarizes the various types of money we discussed and gives a short overview of the advantages and drawbacks of each stage.

2.2 What Roles Does Money Serve?

The key role of money is to facilitate trade. Voluntary trade means that each party prefers to receive the goods that the other party has rather than retain the goods that they were originally holding. Therefore, such a trade improves the well-being of the parties trading. However, as we saw in our historical overview, there are important frictions that limit trade or make it more difficult. Money is an important innovation in that it alleviates some of those frictions. The adoption of a given type of money will depend on

[10] In our brief historical overview, we ignore a number of inventions and novel institutions, most importantly banks. For an overview of the evolution and the role of the banking system, see e.g., Ferguson (2008).

[11] See Sheck (2008).

Table 2.1 A brief overview of the major innovations in the history of money

Money	*Time*	*Positive attributes*	*Negative attributes*
Food based (salt, barley, cacao)	5,000 BC	Easily divisible units (e.g., by weight)	Difficult to transport, perishable (eaten by animals)
Cowry shells, dog teeth, whale teeth	1200 BC	Longer lasting, easier to store	Non-uniform units (naturally occurring in different shapes, sizes and color)
Metal	3,000 BC	Long lasting, easier to store, easily divisible units (e.g., by weight)	Non-uniform due to varying purity; heavy
Metal coins	Seventh century BC	Uniform units (two coins equal), long lasting	Heavy
Paper money	Eighth century AD	Uniform units, mimicked divisibility of units (different denomination), easier to carry	Easy to counterfeit
Electronic money	Twentieth century AD	Uniform units, divisibility of units, even easier to carry	Easy to copy

how well the money's attributes satisfy consumers' economic needs. We discussed a number of such attributes (e.g., divisibility, ease of storage, and transport) in our overview and in its summary table. We now discuss them more systematically.

Economists often use the following three-part definition of money: (1) unit of account (2) medium of exchange (3) store of value. This definition means that two people can agree how much a good is worth in terms of money (that's part (1)); people accept the money when they are selling the good, because they believe it will be accepted elsewhere when you want to exchange it for a good you want to buy (part (2)); and money will not lose its value drastically between the time you get it and the time you spend it to buy something else (part (3)).

These three characteristics make it possible for money to facilitate trade. Each of these dimensions is important. If we know that even one is missing, we would probably not accept a given kind of money in a transaction.

There are, however, some issues with this definition. First of all, it is somewhat circular. In essence, it says that money is something that is being used as money. In this sense, it just describes an equilibrium. What it cannot do is tell us whether a can of mackerel or a Zimbabwean dollar is money. Moreover, the definition sounds like three yes-or-no questions, suggesting that if you answer "yes" three times, what you are evaluating is money. That's not the case.

For example, there is nothing that could serve as medium of exchange in *all* transactions, and nothing that could potentially store value for *ever*.[12] If we'd push for such extreme interpretation, suddenly perfectly good currencies would not satisfy the definition. Take the euro or the Swedish krona or the Polish zloty. Are they good store of value for the next 300 years? That is doubtful. Similarly, the Confederate dollar was money when it was used, but turned out not to be a good store of value—it became worthless after the Civil War. The currency needs to store the value for long enough that the person who gets the currency can reasonably believe they can spend it (few days, weeks, months… the definition is purposely a bit vague on the details here). Otherwise it just would not be a good means of exchange.

Moreover, this definition is meant to apply for a particular environment, for instance, a geographic area. Consider, for example, the Swedish krona. Few people would deny that the krona is money. It certainly satisfies the textbook definition, serving as a unit of account, a store of value, and means of exchange—with one qualification. You can easily transact in Swedish kronas in Sweden, but they may not be generally accepted elsewhere. You are very unlikely to be able to use them in a corner store in the United States.

We see that the textbook definition has an important drawback—it does not state the boundaries, does not define the environment for which it should apply. We cannot apply it universally, as that would make it completely vacuous. For example, even the US dollar, the most global of currencies we have is not accepted *everywhere*. Abroad, one might be able

[12] Notice that "store of value" in the definition does not mean we need money for saving. For saving, we can "invest in something" instead. In fact, money can be inconvenient to keep as savings exactly for the same reason it is money. If it is handy, easy to carry and exchange, then it is also easy to steal. This is why real estate, while inconvenient as money, is more convenient as savings.

to exchange it for the local currency, but not all stores and institutions would accept US dollars directly as a means of payment.

Thus, there is a whole spectrum of how broadly or narrowly this definition applies. In fact, we would argue that some innovations deserve to be called money even though their scope is limited to a few particular transaction types. As we will see, many digital currencies operate with such restrictions, being limited to a particular type of (digital) environment and to only some specific goods you can transact or use within that environment, for example, a sword for your avatar in the multiplayer online game Word of Warcraft. Purists might argue that this disqualifies such digital currencies as "money"—after all they are not a *generally* accepted means of exchange for all, or even most transactions. But then how does it differ from the Swedish krona?

Money should facilitate trade. It may facilitate trade in some geographic area, or only a specific kind of trade. The more limited the trade it can facilitate, the more limited the currency. At some point one can say it is so limited it is no longer a currency. Unfortunately, deciding where that point lies could easily become just an issue of semantics, particularly in an area so new, dynamic, and full of borderline cases as digital currencies.

Given the limits to the textbook definition of money—limiting a currency to geographic region or transaction type—it is easy to see how several different types of money could coexist at the same time. Such a situation has occurred multiple times in the past, as we discuss later.

What Makes Good Money?

Importantly, these limitations do not detract from the incredible usefulness of the definition. Working with this definition, and analyzing the traits that money need to exhibit, has allowed economists to explain why some goods are more fit to be used as money than others. For example, barley is a good unit of account, because it is divisible. But it is not durable and could lose value between one transaction and another; thus it is not a very good store of value. Houses are inconvenient money for different reasons. Even though they are very durable, they are hardly divisible and often incomparable, making them a poor unit of account. It is also cumbersome to exchange ownership of a house—at least, harder than to hand over pieces of metal. So real estate is also a poor medium of

exchange. This is why handy goods that are small enough to carry around and to pass to another person serve the function better.

We can see how the different attributes of different types of money relate to how well each of the three functions is fulfilled. Whether the units are uniform or non-uniform affects the unit of account function. The same uncertainty of whether fish is worth three or four dog teeth, depending on the quality of teeth, makes it hard to assess and compare value of different goods systematically. It may increase the need for bargaining and it makes transactions more time consuming. Thus, such goods do not facilitate trade as well as otherwise similar goods that are uniform across units. On this dimension, barley may be better than dog teeth are. And since barley from different fields may have slightly different qualities, the coins and banknotes that we use today are better than barley.

Similarly, other attributes influence how well a potential currency does as a store of value. Goods that are long lasting and easy to store safely do better as currencies. To take an extreme example, a radioactive element with a short half-life would make for a very poor currency (although, admittedly, its failure as a store of value may not be the biggest problem with it).[13]

Other attributes influence the role of a good as a medium of exchange. Clearly, a well-performing medium of exchange should be easily divisible. Some trades may not be possible if there are no sufficient denomination. Goods that are light and easy to carry do well as medium of exchange: carrying around heavy unwieldy pieces of metal is inconvenient, which makes it tempting to leave such money at home, which in turn may make you miss many opportunities to transact.[14] A good medium of exchange is also not too susceptible to fraud, that is, it is difficult to falsify or duplicate. Scarcity matters for both medium of exchange and store of value. If there is abundance of a particular good, and it is easy to get it in unlimited quantities (e.g., sand on a beach), this good would not make good money. Why would a seller give up a good for sand, if he could easily get the sand and keep the good? To be scarce, money needs to be costly to produce—mine, collect, or grow. For example, metals that function well

[13] For a fascinating overview of how various elements would do as money, explaining why gold is uniquely suited for that role, see Planet Money (2011).

[14] Sweden used copper as money. Because the metal is quite common, you needed a lot of it to transact. Eventually they were issuing 15 kg lumps of copper as money—surely difficult to carry around.

as money—gold and silver—are costly to mine. It was not so much the case with foodstuff money, like barley. Nonetheless, it could still function as money because it did not last long. It was consumed or perished otherwise, and the supply of foodstuff money needed to be replenished every year just to keep the same level. For metal, which is durable and lasts for centuries, to be scarce enough to be money, it needs to be more costly to produce, so that only a small amount is added every year. If as much gold were added every year as barley, gold would quickly lose its value.

The durability of metal also provided a more stable money supply. Barley harvest may be more or less abundant every year. And as supply of money fluctuates, so will the prices. In a year of a good harvest, there is a lot of barley everywhere and the prices of non-barley goods increase. Unstable (that is, changing and unmanaged) supply leads to a greater variability in prices. Such variability intensifies uncertainty, which in turn creates frictions in trade. It makes metal, with its more stable supply, more preferred as money. Of course, even metal money supply may experience large fluctuations. The primary example is the discovery of Americas, which brought large amounts of gold and silver to the European economy.

For most of the history of money, people could choose whether to "produce" money or produce goods and services that could be exchanged for money. Growing barley, mining metal, or looking for cowry shells is how one would produce money directly. But for that, on would need to make a choice to grow more barley instead of grazing cows, or abandon their farm to look for gold in California's rivers. Such choice was no longer possible with the introduction of paper money. Paper money was cheap to produce, and its scarcity came from state regulation in the form of strong constraints on who could produce money and how much. Thus, scarcity of paper money was imposed artificially, while scarcity of earlier money resulted from the cost of their production. As we will see later, the issue of scarcity is very important for digital currency schemes, as digital money could sometimes be made "with a click of a mouse." This issue was especially challenging for decentralized digital money systems.

Table 2.2 sums up these arguments, highlighting various attributes that support the three roles of money. In our historical overview, we saw how these roles and attributes influenced the evolution of money and led to gradual improvements in how we transact.

Table 2.2 Attributes supporting the critical roles of money

Role of money	*Attributes supporting the role*
Unit of account	Uniform units
Store of value	Long lasting, easy to store securely, scarcity
Medium of exchange	Easily divisible, uniform units, light and easy to carry, trustworthy (less susceptible to fraud), scarce

Transaction costs

The importance of the three roles of money, and the attributes that support them, is related to transaction costs. Broadly speaking, money facilitates trades by lowering transaction costs. And more transaction costs can be overcome when money satisfies its three roles well.

All transactions have some element of costs inherent in them. The costs may come from many drivers. Perhaps the most obvious one is the time needed to conduct a transaction. We saw the importance of this cost already in the earliest human communities, as it was one of the most important costs of barter: you may need to spend a long time to find somebody willing to trade something you have for something you want. To a lesser degree, time costs made un-minted pieces of metal inferior to later types of money: you needed to spend time weighing a piece of metal or dividing it into smaller pieces. Another type of cost is related to the effort in changing the ownership of the means of exchange. For example, money that is particularly heavy or difficult to transport would be costly to deliver to the seller.

Other important costs are the mental costs, for example, having to conduct relatively more complicated arithmetic to complete a transaction that uses multiple different units of a currency, or multiple different currencies. A related cost is the probability of making a mistake for example, in deciding how much change to give back, or in distinguishing differing qualities in dog teeth or pieces of metal that might influence their value.

Besides these costs, there exist transaction costs that are more indirect. After a completed transaction, the seller may need to secure the money he has just obtained, which, depending on the type of money, may be costly. The obvious example here is the protection from theft, for example, hiring guards when transferring money, building safes for storing metal, and so

on. Less obvious examples, more relevant for commodities-based currencies, are the need for protection from the elements and vermin, or the need to build large warehouses to store your money; both are quite important when the money is, for example, barley.

Finally, lost opportunities—foregone transactions that did not occur—are another type of transaction cost. Money that does not satisfy its three roles well may not be able to facilitate as many transactions, and each transaction that does not happen is a loss to the potential buyer and seller, and to the overall economy. The attributes of the good serving as money may contribute to the loss of transactions, for example high weight of or lack of familiarity with a particular metal. Because of these attributes, potential trading partners may view the transaction as too costly to conduct, or perhaps too risky, and decide not to go ahead with it. Transactions may be lost also when units of the currency are not sufficiently divisible. For example, if a particular fish is worth 4.25 dog teeth to the seller and 4.75 teeth to the buyer, their trade would be beneficial for both sides, but it will not occur because dog teeth are not divisible. A transaction might be conducted for 4 teeth or perhaps for 5 teeth—but it won't, as either option would make one of the parties strictly worse off than not transacting at all.

The transaction costs argument helps us understand why gold has been a long-time winner in the money arena. Gold is durable, divisible, and can be weighted for a uniform unit of account. Moreover, gold is valuable, even culturally, because it does not change its appearance over time.

Trust and Counterfeiting

Gold helps us highlight a particular attribute of money that will become important with digital currencies: trust. Money should be a good store of value, and scarcity is often thought to guarantee the value over time. It is also relatively more difficult to falsify—or, at least, tools such as touchstone were developed to check for the purity of gold.

Trusting that a currency is genuine is an important prerequisite for conducting a transaction. Although nowadays we usually think about counterfeiting in the context of paper money, this nefarious procedure is much older than that. For example, metal coins were often "clipped" making them of lower weight than they should according to the stamp. To prevent debasement, coin edges were stamped or rimmed, making it easier for users to identify whether pieces of metal were cut from a

coin, changing its weight and its value. Nowadays coins' value no longer comes from their weight. Nonetheless, many contemporary coins have rimmed edges, due to this legacy. In another type of counterfeiting, metal coins or unminted metal pieces could contain a lesser-valued metal inside, obscured by the correct metal outside. Imagine, for example, a copper core covered in a silver coating to imitate a silver coin. Human ingenuity is limitless. Even commodity-based money was falsified. Consider cacao, used in the Aztec Empire as money. Counterfeiters falsified that currency by filling an empty cacao husk with mud and sealing it.[15]

Counterfeiting considerations are particularly important in the context of digital currencies. Digital technology makes it very easy and cheap to make perfect copies of digitally stored information: files, code, passwords, addresses, and so on. In the music industry, it resulted in large-scale piracy, which changed how this industry operates. In the context of money, it gives rise to the so-called double-spending issue.

In the next chapters, we will analyze the various roles of money in the context of digital currencies. We will then see that many of the attributes are as important to traditional (physical) and to digital currencies. We will see that digital money may have significant advantages when it comes to facilitating trade, making it cheaper and faster. We will also see that fraud, and hence the lack of trust, has been a particular challenge for attempts to create money in the digital world.

2.3 Competing Money

Most of us are used to one particular type of money (say, US dollars) and we think of that "the money" as just being there. There is nothing wrong with this perception; in most places, at a given time and place just one particular currency is in use. But as with any other product, money competes with other money. If we look closely, we will see this competition all the time. In the historical context, silver competed with barley, metal coins competed with unminted metal, and paper money competed with gold. Interestingly, multiple competing currencies often coexisted, if only for some time. Venetian ducats and Florence's florins competed with other coins throughout medieval Europe, and now the euro and the US dollar compete in international transactions. In fact,

[15] See Weatherford (1997).

without competition there would be no change—a new currency or a new form of money is introduced into an economy that typically already has an incumbent currency. The new innovation can only survive, and perhaps eventually win widespread adoption, if it can successfully compete with the incumbents. But then, what determines the outcome of such competition?

2.3.1 Coexistence Is Costly

There are clear costs to having multiple currencies within an economy. We can divide these costs into two broad categories: cognitive costs and costs of exchange.

The cognitive costs arise from mental hardship of having to compare prices and values quoted in various currencies. One needs to not only compare different units when deciding whether to buy something but possibly also perform some mental arithmetic when selecting the banknotes and coins to pay for the purchase, or when accepting change from the purchase. Consider, for example, the coinage system in England. That system historically included farthings, pennies, shillings, crowns, pounds, and guineas, some made of different metals, and thus changing value to one another. Finally, the relative value of these different units was fixed in 1717. For example, the value of a guinea had fluctuated between 20 and 30 shillings, before being fixed at 21 shillings in 1717. A pound contained 20 shillings; so a guinea was worth 1 pound and 1 shilling. A shilling contained 12 pence and each penny contained 4 farthings (and, in earlier times, it varied between 8 and 4 farthings to a penny). Crown was a quarter of a pound.

Other European countries also used multiple units. For example, the pre-revolutionary France had a system of currency that rivaled the English one in terms of its complexity. The central unit of the system was the louis d'or, which consisted of 10 livres. Each livre consisted of 20 sols. Each sol consisted of 12 deniers. And those were just gold coins. Among silver ones, 60 sous constituted 1 silver ecu. The relative value of gold and silver coins was changing with time. Such multiplicity created frictions. The local population must have been used to this mélange. Nonetheless, one suspects that this multiplicity of types of coinage created much scope for mistakes and confusion. With similarly complicated and incompatible systems in other countries, it made international trade more confusing. Eventually, such frictions were resolved by adopting the metric system

which strongly relies on decimalization. Decimalization of coinage started with the United States and France in the late eighteenth century. The UK had been a holdout in its long-standing refusal to adopt the decimal system in their currency. The system with pound as a unit and 100 (new) pennies to a pound—dropping other units, like guinea and farthings—was only introduced in 1971. Decimalization of currency decreased the mental cost of handling money: if one operates in decimal system, it is much easier to add, subtract, or multiply values expressed in currencies quoted in the base of 100 (as opposed to, say, 21, the number of shillings in a guinea).

Technology can help diminish these costs, although arguably not eliminate them. For example, cellphones and widespread Internet coverage make it easy to convert prices quoted in a foreign currency into your home currency. Still, there is, and likely always will be, some inconvenience in, say, having to turn to your cellphone every time you want to buy something. Moreover, even if referring to your cellphone is hassle-free, it does not preclude the second large category of costs: costs of exchange.

In economies that use multiple different currencies, people bear the cost of having to exchange one currency for another. This cost cannot be avoided at the level of the overall economy: even if you decide to only ever accept and spend one type of currency, some of the parties you transact with will need to exchange your favored currency for the currency of choice of their other customers or suppliers.

To better illustrate the costs of multiple different currencies circulating in an economy, let's consider the state banking era of the United States in the period between 1786 and 1863. In those early days of the country the U.S. government minted coins, but did not issue paper money. The reason for this setup was that government-printed money was subject to controversy after the overissuance of Continentals during the War of Independence.

Even though the U.S. government refrained from issuing paper currency, private banks printed their own paper money, eventually supplying the market with a plethora of various banknotes. The issuing private banks were established based on individual states' legislations, and virtually every private bank issued its own notes. The scale of this phenomenon is illustrated by the fact that in 1860 there were over 1,500 banks in the United States, out of which 54 were in just New York City.

The banks were not allowed to simply print money at will. By the requirement of the legislature, the notes they issued had to be backed by assets, and the issuing bank had the obligation to redeem the notes for specie, that is, metal coins. A failure to exchange the notes brought for redemption into specie was a serious offense and it could be a cause for the bank's failures. On average, 0.5% of banks failed annually, although there were years when even 5% of banks failed.

With thousands of different types of banknotes circulating in the economy, not all notes were treated equally. For example, it quickly became clear that a five dollar note from one bank could be worth less than a five dollar note from another bank. These discounts made the exchange of banknotes and trade more costly.

The reason for different valuation of notes often related to the difficulty and risk of successful redemption of the note for specie. To redeem the note, one had to go to the bank that issued it. This may have been easy for your local bank, but would have been difficult and perhaps too expensive if you had a note issued by a bank far away. If you still undertook the journey, and if were particularly unlucky, you might have found that the bank you were going to had failed by the time you got there. Indeed, researchers found that the discounts varied geographically, and discounts were generally lower for banks that were local and, hence, more known to people living in a given area.[16]

The discount also captured the risk of a bank failing. If such a risk was high, it was less certain that the banknote could be redeemed. Failing banks either would not redeem notes at all, or redeem them at a fraction of the face value. Thus, accepting notes from some banks was considered riskier than accepting other banks' notes. It may have come from general knowledge that a particular bank was in trouble, but also from lack of familiarity with the bank. Somebody who lived in Philadelphia may have had less information about Boston banks, and may have been less willing to accept banknotes issued by those banks. This was another reason why the notes from far away banks traded at a larger discount.

Uncertainty about the value of a banknote ties to another phenomenon: forgery. Counterfeiting was rampant. With the multitude of note designs it was difficult to keep track of what a genuine note of a particular bank should look like. Again, it was more likely that

[16] See Weber (2014).

banknotes from afar were counterfeits, as people were less familiar with their design. More colorfully, forgers would sometimes make up entire banks and banknotes issued by these (fictional) banks. In the environment with hundreds of different issuers, forgers sometimes managed to get away with this ploy, and ultimately it contributed to people's general aversion to less popular banknotes, or banknotes from geographically distant locations.

You can imagine that most people were simply unable to keep track of all these issues and nuances. Not surprisingly, brokers appeared who were willing to accept various banknotes and exchange them for others—for a price. The brokers in many cities would publish weekly, biweekly, or monthly "counterfeit detectors" or "bank note reporters"—publications listing known counterfeits and often quoting discounts for trading genuine notes of different banks. In those publications merchants would find advice such as "better refuse all 5s" from Webster Bank of Boston, Massachusetts, or "beware of all denominations of the old fraudulent Bank of this name" for New York Exchange Bank.[17] These reporters were available to the public—again, for a price. But even if you had one, consulting it was time consuming for merchants, and others who were using them.

Overall, the costs of having this multitude of banknotes were high. They included both cognitive and economic costs. The latter included the direct costs of conducting transactions (e.g., having to buy a currency reporter) and the costs of bearing the extra risk and uncertainty when dealing with various banknotes. All this has created frictions in trade and a burden to the overall economy.[18]

The desire to avoid these costs is an important driver of competition among currencies and may eventually push the economy to one generally adopted currency. As it turns out, there is also another powerful incentive operating in the same direction: network effects.

[17] Thompston's Bank Note and Commercial Reporter in New York, January 1, 1854, cited after Weber (2014).

[18] In order to eliminate these costly frictions, US passed National Banking Act that took hold in 1863. The aim of the Act was on-par acceptance of banknotes throughout the country. It was achieved with a clearinghouse operations and insurance schemes.

2.3.2 Network Effects

Competition between currencies is different from competition between most goods, and one aspect plays a key role here: money exhibits what in economics is termed "network effects." Simply put, an object is more useful as money if other people are using it as money as well.

Network effects were first recognized in economics in the 1980's.[19] To use the most classic example, consider the telephone network. There is no use to own a telephone, if you own the only one. The value of a telephone increases as more people buy phones, i.e., there are more phones in the network.

Over the past few decades, studying network effects became a vibrant sub-field of economics. Tools that economists developed to study networks have been used to analyze, explain, and understand a variety of modern technologies: videogame consoles, computers, or smartphones. The applications are particularly relevant in the context of communication technologies. In fact, it has been observed that what has been named "network effects" do not need a physical network. There is no need for wires like in the telephone network for network effects to occur.

The network effects logic readily applies to money. Suppose you want to introduce a new form of money. Initially, you are the only one who recognizes and accepts that money, making it very difficult to persuade someone else to adopt it as well. After all, if he does, he will initially have only you to trade with. Things are easier if there is already a larger part of the society, hopefully including both potential buyers and sellers, who stand ready to use the currency.

With network effects, we often see "winner takes all" dynamics. If two networks are similar but one is larger, the larger one will be more attractive to the new users. Users from the smaller network may also prefer to switch to the larger network. The larger will grow even larger, while the smaller may even disappear. Thus, the winner takes the whole market. Often such market is efficient, as all users may take advantage of maximal network effect, as they benefit from having access to everyone on the same network. Because of that, economic research often finds that it is socially optimal when we all use the same technology that generates network effects.

[19] See Katz and Shapiro (1985), Rolfs (1974).

We frequently see such winner take all dynamics in the context of money. As with other technologies that generate network effects, money accepted by a larger number of people is more useful than money used by a few. And since a currency is more useful when more people adopt it, the benefit is maximized when everybody uses the same currency.

In our earlier historical overview, we discussed the appearance of coins in Lydia in the seventh century BC. There were good reasons why coins were a superior technology to unminted metal—for example, the coins with the same mark were uniform, they were all worth the same, and everyone knew what they were worth. They saved time on weighing and decreased probability of cheating. Thus, when two trading parties could use coins or unminted metal, both preferred to use coins. Moreover, the seller knew that he would have an easier time using coins rather than unminted metal in future transactions, so he was more willing to accept them. And as more people used coins, fewer people wanted to use unminted metal. That is, as coins became more popular, their appeal grew and it further increased their popularity. With time, coins took over the market for most transactions. Unminted metal was used when coins were not available or when value of a transaction was very large and one slab of metal was handier than many coins were.

The Renaissance gives us another example of the winner take all dynamics in money. During the Renaissance, Italian banking—especially Florentine and Venetian—spread throughout Europe, making the currencies of Florence (florin) and of Venice (ducat) the currencies of choice even in places far away from Italy. With credit from those Italian banking houses, many trades were conducted in those currencies and people became increasingly familiar with them. When merchants had a chance to conduct trade in florins and ducats or some other coins, they preferred florins and ducats. Thus, florins and ducats were gaining popularity, becoming the dominant currencies of Europe, and pushing out other currencies.

Our final example is that of the Maria Theresa thaler. The thaler (a name from which the word "dollar" is derived) was introduced in 1773 in honor of the Austrian empress, the wife of Holy Roman Emperor Francis I. It rapidly became very popular, especially in North Africa and in the Middle East.[20] People became reluctant to use any other currency. The

[20] See Weatherford (1997).

reason why they preferred Maria Theresa thalers is precisely the network effect: they preferred the thalers because they knew that everyone else would also prefer to trade using Maria Theresa thalers, and may not be as inclined to trade using other potential coins. This dynamics reinforced the popularity of Maria Theresa thalers in the region, pushing other coins out.

Maria Theresa died in 1780, but the coin continued to be minted. It was an unusual practice to mint coins with an image of a deceased ruler, so all the coins minted after Maria Theresa's death bore the date 1780. They kept being minted after Napoleon abolished the Holy Roman Empire in 1805. And after the Austro-Hungarian Empire disintegrated following World War I, the Austrian Republic continued to mint them until the Anschluss by Hitler in 1937. Italy minted Maria Theresa thalers in the late 1930s for the use in the conquered territory of Abyssinia (today's Ethiopia). Tellingly, Mussolini's government decided to supply the thalers because the local population in Abyssinia refused to accept substitutes. Maria Theresa thalers were more familiar and trusted, and the power of the "winner-take-all" dynamics was so strong that it was difficult for modern currencies to be successfully introduced into that economy. This dynamics was not limited to Abyssinia: the thaler was minted in mints from Bombay and Brussels to Utrecht and Vienna. Even after the World War II, Austria resumed minting the coins in 1956—the last being minted in 1975. The total number of silver Maria Theresa thalers minted between 1780 and 1975 is estimated at about 400 million. Each one is dated 1780.

With network effects pushing the economy toward a single currency, why do we observe prolonged episodes in which multiple currencies are in use, for example, the multitude of banknotes during the state banking era in the United States, described earlier? In the case of the banking era, the reason was the external limit imposed by regulation. The coins that won the market, whether florins, ducats, or Maria Theresa thalers, were minted up to the point when the supply of the coins matched the demand. In contrast, banks under state banking laws were kept small (e.g., they could not merge with each other) and they were limited in the value of banknotes they could issue. The issuance was limited by the banks' capital, which in turn was limited by the law. For some small or sparsely populated areas one bank's supply of banknotes was enough to match the demand. But for most urban areas, the demand for banknotes was much larger than what any one could legally provide. This restriction, and the situation it gave rise to, was detrimental for the economy as a whole and

some standardization was needed. As we will see below, it was a central authority (essentially, new regulation) that solved the problem: The US government forced all banks and citizens to use the US dollar.

2.3.3 The Difficulty of Introducing a New Currency: Excess Inertia

Time and again we see innovation—say a new and promising technology—that has problems wining market share from the incumbent that may be offering a less efficient technology. Network economics allows us to better understand this tug of war between popularity and ease of use. This interplay, as identified in the economic literature, is one of the characteristic features that we should expect in environments with network effects. Such environments are often too slow in adapting new technology, and they sometimes fail to adopt it altogether even though it would have been beneficial to do so. Economists call this "excess inertia."[21]

In our historical overview, we saw innovations that were seamlessly introduced into the economy and that eventually won widespread popularity. For example, coins were quickly adopted and they eventually crowded out the prior incumbent, unminted metal pieces. However, other innovations faced major frictions, slowing down adoption or making it outright impossible.

Such adoption friction was present in the case of paper money. Paper money is a better technology, in terms of convenience, than metal money. For example, it is easier to transport. Yet, it took a long time for the Western world to embrace it. In contrast, China adopted paper money much earlier, because of the direct enforcement of this innovation by the state.

Similarly, credit cards are more convenient to use than cash, especially for large-value transactions. They are appealing to customers because they are lighter and safer than cash and eliminate the need to worry about change. Their appeal is somewhat more limited for merchants, who need to pay additional fees to be able to accept credit cards. Nonetheless, for large-value transactions the benefit of increased security may outweigh the cost, because for example the merchant may avoid carrying large amount

[21] The network effects literature also recognizes "excess momentum" where people adopt a worse technology too early because they expect everyone else to do so as well. But this is unlikely to occur in the context of currency. People tend to be very conservative when it comes to innovations related to money.

of cash to the bank. Moreover, by accepting credit cards, the merchants avoid the risk that the trade would not happen because the customer does not have enough cash on him or her.

Indeed credit cards became very popular, at least by the turn of the century. However, the initial adoption was not very brisk. Despite the advantages of the technology, it was more of a push of the credit card companies than a pull of the customers. There was a lot of mistrust, both on the side of customers and on the side of merchants. To counter that inertia, credit card companies put a lot of effort in educating people and encouraging the use of the system. For example, they give rewards for using credit cards, and they advertise their fraud protection plans.

Credit card companies do not issue cards and manage payments only for the social good and the benefit of the market. They are concerned with their own profit. But one could easily imagine that without the active role of credit card companies the market would stick for longer to the traditional but less efficient use of large amounts of cash. Alternatively, the new technology could have fizzled out because each side would worry that the new payment system would not gain enough traction with the other side. Nowadays we can point to the great convenience of using credit cards online, and think that the benefit of adoption is clear. But credit cards would probably not be used online if they had not been adopted earlier, for brick-and-mortar transactions.

From the examples above, we see that sometimes the ease of use is the prevailing force and the new technology is smoothly adopted, like coins. Sometimes it is adopted with resistance and frictions due to excess inertia, as with paper money and credit cards. And it is possible that sometimes it is not adopted at all. We simply do not observe a failed potential entrant. For instance, it may be that the popularity of Maria Theresa thaler hindered adoption of some better forms of currency.

Our final example of excess inertia comes from the United States in the 1860s. As described earlier, until 1863 all the banknotes in circulation were provided by private banks under individual state banking laws. Counterfeiting was rampant and occasionally banks were failing, rendering notes useless, or redeemed at very high discount. In 1863 banks started issuing notes under a new legislation, the National Banking Act. Those so called “national banks” were still private banks, usually with a single brick-and-mortar location. But the notes they issued were of a distinct, uniform design, which made it easier to control for counterfeits.

Moreover, the national banknotes were insured, which meant that even if the issuing bank failed, the notes would be fully redeemed for specie.

Given that national banks' notes carried less risk than state banks' notes, they were more reliable money. When passing the law, the government expected that with such an advantage, national banknotes will naturally become widely accepted, rendering state banknotes obsolete. However, after two years there was no visible decline in the use of state banknotes. Since bank failures occurred only occasionally, people may have considered the risk a natural part of the transaction costs. At the same time, they may have been distrustful of the unfamiliar design, and may not have been fully aware of the benefits of the national banknotes. State banknotes were more familiar, and people knew they were accepted in their immediate environment. So state banknotes kept being accepted because everyone expected they'd be accepted.

The government effectively put an end to state banknotes by putting a ten percent tax on banks paying the state banknotes out over the counter, even if they were the bank's own notes.[22] This finally ended the era of state banking.

2.3.4 Coexistence of Various Currencies

Despite the winner-take-all dynamics and despite excess inertia, sometimes different forms of money, different currencies can coexist in the economy. This happens when the different currencies serve different purposes.

We have the first records of silver used as money from ancient Mesopotamia. It replaced an older type of money—barley. Metal held a higher value than barley: a piece of silver was worth more than the same volume of barley.[23] This is why silver was more convenient for transactions involving large values and longer distances (e.g., a shipload of products). For everyday local exchanges of much smaller values, metals were too valuable. Those trades were still conducted using barley.

[22] See Weber (2014).

[23] Originally a piece of silver was worth the same as equivalent weight of barley. The value of silver was counted in shekels. And the word "shekel" is derived from "weight of barley."

Thus, even though metal was handier and was adopted throughout Mesopotamia, winner-take-all dynamics have not led to metal money pushing out older barley money completely.

Similarly, the introduction of coins has not completely eliminated the use of unminted metal in transactions. That was the case especially for high value transactions, where a large number of standardized coins would be unhandy. Different transactions have different "needs," and different currencies may coexist if they serve those different needs better. There are still costs of parallel money—exchanging barley for metal and vice versa, but the benefits of matching functionality to needs may be worth the costs. The two types of money serve their respective purposes better than having only one type.

We can also think of contemporary banknotes and coins as two different kinds of currency that coexist because they serve different purposes. We tend to use banknotes and coins for different types of transactions. Typically, we use coins for small value transactions and banknotes for large value ones. Sure, there is overlap, but if we only had banknotes or only coins, trades would be more laborious. And for their respective roles, the two types utilize optimal technology.[24] Banknotes of very small denominations that circulate very frequently would wear out too quickly. Coins are more durable, but they are heavier than banknotes. Using many coins, even of higher denominations, for large-value transactions would be less handy than using bills of the same denominations. Customers would need to carry fewer coins if more denominations were available. For that to work, merchants would need to have all denominations always available, and with larger number of denominations, they would need to tie up more of their capital just to have change ready. The transaction costs would also increase because one would need to search for coins among more denominations.

These different roles that coins and banknotes play were apparent from the time banknotes were introduced. For example, some of the first banknotes issued by the Bank of England in the eighteenth century were the ten-pound and twenty-pound notes. These quantities, equivalent to roughly a thousand dollars nowadays, limited the use of the banknotes to

[24] In some cases, however, cultural legacy may prevail despite sub-optimality. Several arguments speak to the point that $1 US bill is not optimal, and should be replaced by a coin for durability and cost of handling. See The Economist (2013).

the richest strata of society. Not surprisingly, they were used almost exclusively for large-value business transactions, and were particularly popular among the financial elites of the City of London.[25]

Overall, we can summarize the competitive forces as follows. There are costs to multiple currencies, including not only the cognitive costs but also the cost of exchange. Different currencies are available—some may be better or worse than other for a particular purpose, and some may be equivalent. People are willing to use multiple currencies and bear the cost of compatibility and exchange if the currencies serve different purposes and if each is better for its purpose than others. But people would rather use one currency for a given purpose: network effects matter for currencies. Network effects tilt the economy towards winner-take-all outcomes, where a single currency accounts for all transactions in the economy. In such cases, the incumbent currency may hinder competition, with inertia keeping people from adopting new (or multiple) currencies that could improve their well-being.

Our overview of the history of money brings us to modern times and digital currencies. The background we covered in this chapter will give us a better understanding of various technological innovations in digital money and will help us highlight the similarities between them and earlier stages in the evolution of money.

[25] See Weatherford (1997).

CHAPTER 3

Platform-Based Currencies

In the early 2000s, many large Internet companies have introduced their own digital currencies. Most of these companies run large platforms that span media, entertainment, and e-commerce. The market has seen Amazon Coin, Facebook Credits, QQ coin, Microsoft Points, Reddit gold to only name a few. This is on top of many video games, gaming platforms, and metaverses that have introduced their own currencies, e.g., World of Warcraft gold, Second Life's Linden dollars, or Eve Online's ISK among many others. For interactive video games it has become standard practice to have a dedicated, game-specific currency, which in some sense allows the game publisher to claim property rights over the digital items in circulation within the game. The explosion of the so-called "freemium business model" by game publishers, where anybody can play for free but there are options for purchasing digital goods (weapons, cloths, etc.) has made this even more important since all revenues come from in-game purchases. For example, it is estimated that Epic Game's blockbuster video game, Fortnite, earned over $5 billion in its first year, all from in-game purchases from its vast user base counting hundreds of millions of players. Strikingly, in-game spending (of so-called V-Bucks) does not provide any advantage to the player in Fortnite's Battle Royale (e.g., better weapons or protection); it only makes the player look "cooler." While Fortnite's success is quite exceptional, its freemium business model is quite common. It is not surprising therefore that total

H. Halaburda et al., *Beyond Bitcoin*,
https://doi.org/10.1007/978-3-030-88931-9_3

in-game spending by video game players across the world has steadily grown to reach some $129 billion per year by 2020.

All of these digital currencies have been introduced by online platforms businesses that, in one way or another, help interactions between their large and diverse members (buyers and sellers, game players or, simply, people who want to exchange pictures and messages with one another). These interactions often involve some form of trade that may be helped by a special (custom-built) currency that online platforms provide for the convenience of their members. It is important to see that in all these cases the currency is entirely controlled by the platform, which can set all of its features and properties. In this chapter, we review a few such "centrally controlled" currencies to understand the key drivers of their design and the rules governing their use.

Special-purpose money centrally introduced and controlled by various organizations from commercial entities to local or national government organizations are not entirely new. Casino chips and Monopoly money have been around for a century. Also, while they are rarely called currencies, the world has been quite used to airline miles redeemable for future flights, hotel reservations, or car rentals (often purchased for friends and relatives). And airline miles are just one example of a family of loyalty programs in millions of stores or for a multitude of products and services. Governments have regularly introduced actual currencies restricted to specific social groups, geographic regions, or product categories. For example, there are still a number of local currencies functioning in different regions of the United States, e.g., Ithaca Hours in upstate New York or Berkshares in western Massachusetts. Food stamps are another example of special-purpose money: they are essentially a payment system restricted for use by the poor and only for certain products. Even a mortgage can be seen as a restricted currency.

What has changed, however, is that the digital era represents vast new opportunities and challenges for the introduction and use of special-purpose currencies. Firstly, the digital nature of these currencies provides endless opportunities for the design of new features adapted to the specific needs of the business introducing them. Also, besides the multiplicity of features, the digital era also makes it much more cost effective to monitor and restrict the use of the currency. These varied needs and opportunities explain much of the differences between today's platform-based currencies. Yet, in many cases, early experimentation with the digital currency made the businesses realize that a special currency does not help their

customers (and their bottom line) and lead these businesses to abandon the currency. Most importantly, however, many of the recently introduced digital currencies are *global*. The organizations offering them are often large platforms, spanning across many (in fact, typically most) countries. As such, these currencies may have a global impact. And this fact did not escape the attention of policy makers and economic commentators. Matthew Yglesias (2012), mentioned in the Introduction, worried about Facebook Credits taking on established national currencies and were echoed by economists who saw the coincidence of these currency introductions with the rise of developed countries national debt after the financial crisis particularly threatening for national currencies. Similar concerns were expressed when Amazon introduced Amazon Coins in 2013. Experts saw a potential for these currencies to challenge central banks' monopoly on issuing money. Besides the historical context of the financial crisis, these concerns were certainly also fueled by the fact that Facebook and Amazon are large platforms with broad international reach and very large customer bases with billions of users. For example, Facebook's 2019 announcement about the planned introduction of Libra, a digital currency to be used by its members, has met broad international resistance by regulators. Indeed, it is often reminded that with its size, if Facebook and the other platforms it owns (WhatsApp and Instagram) were a country, it would be more populous than China and India combined.

With the benefit of hindsight, concerns about platforms' early attempts at introducing digital currencies have largely disappeared and not only because Facebook decided to abandon Facebook Credit (and, the plans about Libra, now called Diem, keep evolving). As we will argue below, these early currencies had no real potential to become widely accepted currencies despite the large size of their patron companies. The main reason is that they were severely limited in their functionality. For example, neither Facebook Credits nor Amazon Coins could be transferred to other users, and they could only be spent on Facebook or Amazon. Amazon Coins had additional restrictions on what they could be spent on—only on selected apps on Amazon Kindle Fire. With such limitations they could not become a means of payment rivaling the dollar, euro, or the yen. Indeed, transferability is necessary (although not sufficient) for a platform-based currency to have a wider impact. Yet, platforms did not impose these limitations to its currencies by mistake. Platform currencies with limited functionality may enhance the strength of the

network effects and as such may be an effective strategic tool in certain stages of platform growth.[1]

This is not true for currencies that indeed aim for full functionality, as Facebook's planned Diem (previously announced as Libra) mentioned earlier, and for which commentators' concerns are well justified. In the past, some platforms introduced currencies with full functionalities that can be freely exchanged for national currencies (e.g., Second Life's Linden dollars can be exchanged back and forth between US dollars). While these currencies had no major influence on national currencies so far, this is largely due to the fact that the underlying platforms failed to grow large enough for such impact. Moreover, even limited local currencies will represent a challenge for regulators who will find it hard to coordinate across national borders to implement regulation. Yet, with the flexibility in design that the digital nature of these currencies makes possible, such regulation might be increasingly necessary.

What drives this design? To understand the larger picture of digital currencies, we need to examine more carefully the incentives of Internet companies when issuing their currencies. Special purpose ("local") currencies have always been introduced with specific objectives in mind. Their design closely reflects these objectives while trying to avoid unintended consequences. This has been the case for non-digital local currencies as well, as we will illustrate below. Coming back to our digital examples, Amazon and Facebook have already grown large before introducing their currencies. They operate according to their specific business models, and their spectacular growth may be an indicator that these business models are successful. We venture the hypothesis that the companies only introduce their currencies if it reinforces their business models. The main insight is that digitization allows for the *design* of currencies to unprecedented extents and companies are designing their currencies by choosing the currency's attributes in such a way as to best match their business models.[2]

[1] See Gans and Halaburda (2015) and Fung and Halaburda (2014).

[2] Notice we are not talking about companies whose main business model is to facilitate payments with regular national currencies, like PayPal, M-Pesa, or Venmo. The key difference is that these payment platforms do not introduce alternative currencies. We will also refrain from analyzing platforms that may facilitate the (sometimes illegal) exchange of special currencies introduced by other platforms, e.g., eBay. They have their own challenges and potential. We will discuss some of them in more detail below.

In what follows, we first review a few classic examples of centrally introduced local currencies and show how their design features reflect the underlying objectives of the organizations that have introduced them. Next, we look at four typical business models by large Internet platforms and analyze how the features of their recently introduced digital currencies reflect these business models. Finally, we discuss the limits to the distinction that can be made between platform-based currencies and government-issued money and the challenges that large-scale digital currencies might represent.

In our analysis of the following examples we will focus on three main attributes, which can be easily set and controlled by the organization introducing the currency. Arguably, these attributes have a major impact on whether the currency can facilitate trade (a currency's core purpose) and in what specific context they can do so.[3] The first such attribute is *acquirability*, or how the currency can be acquired. The designer of the currency can, for example, impose that the currency can only be "earned" with certain specific activities or that it can be "bought" (exchanged) for other currencies or goods. The second feature that we examine is *transferability*, or what are the restrictions (if any) on transferring the currency to others. Typically, the question is whether it can be transferred to (and/or which) other members on the platform. Finally, the third feature, *redeemability*, prescribes what the currency can buy? In particular, of central interest is whether it can be exchanged back for other currency with less restrictions (typically for national currency). In other words, redeemability defines the restrictions on spending the currency. If a currency does not have restrictions in any of the attributes, i.e., can be bought and earned, can be transferred to anyone participating in the system, can be exchanged back for fiat currency and spent on anything within the system, we call such currency *fully equipped*. National currencies can be considered fully equipped currencies, at least within the country whose government issued them. Most digital currencies, however, are typically restricted in one or multiple attributes. Those restrictions are deliberately put in place in order to reinforce the business model of the issuing platform. Let us consider several examples in more detail, starting with some traditional ones, rooted in the non-digital world.

[3] See Gans and Halaburda (2015) and Fung and Halaburda (2014).

3.1 Special Currencies of the Traditional World

As mentioned earlier, the design of money has been around for a while. All kinds of loyalty points with restrictions, food stamps, some of the banking products (e.g., mortgages) are examples of such a design. Below, we will look at three particular examples: Berkshares, food stamps and the mortgage to analyze the design challenges that they have faced in light of their issuer's objectives.

Consider **BerkShares**. These were introduced in 2006 in The Berkshires region of Massachusetts with the intention to help the local population in a touristy area. The presence of tourists increased prices in the area, but not necessarily local wages. Some local businesses got together and agreed to give a discount to the local population—in essence, introducing an effective price discrimination scheme. This was done through BerkShares, essentially a local paper currency. You could get BerkShares at a local bank paying 95 US cents per one BerkShare. But the participating businesses accept them on par with the dollar.

Since they are a paper currency, any restriction on transferring them between local and non-local people would be too costly to enforce. Moreover, when you use BerkShares you do not need to prove that you are a local (or even pretend to be local). One could imagine a requirement that you need to show your driver's license with a local address to use BerkShares. But maybe this would be too burdensome or elicit negative sentiment from tourists, and would slow down transactions. Since you can only use them at participating businesses, BerkShares have restrictions on where one can spend them, but not on who can spend them. Anyone can buy them at a local bank. They aren't advertised, so not many people know about them. But, of course, locals would be more likely to know about them. This has been probably the only barrier for everyone taking advantage of the discount. Surely, if too many tourists would take advantage of BerkShares, one can easily imagine additional restrictions on acquiring and spending BerkShares to take hold. Since those are costly to implement (in additional time and burden it takes to complete transactions), they wouldn't be implemented until there is a need to do so. For digital currencies such restrictions are much less likely to bear these additional transaction costs. They could be incorporated in the design right from the start.

Food stamps is another example of designed money, one, where restrictions on spending are actually in effect. You can only spend them

in particular places, on particular products—only food, and no alcohol or tobacco. The purpose of food stamps is for the government to provide food to families with low income. In such a case, giving those families the same funds in cash would enable the families to spend that money on goods other than food (including drugs or alcohol). This would contradict the purpose of the program. Introducing a distinct currency, with restriction on its use allows the government to achieve its goal—supplementing food to those families in need. Originally, food stamps had the form of paper stamps or coupons, similar to paper currency. They were accepted by participating food stores no matter who was using them. However, that meant that there was no restriction on transferring them (eligible families could pass their food stamps to non-eligible families) and no restrictions on who could spend them. Their use was only restricted by where they could be spent and on what products. Since the late 1990s the paper stamps were phased out and replaced by a debit-card system (Electronic Benefit Transfer—EBT) administered by banks, presumably to save costs. In 2008, the government renamed them Food Stamp Program to Supplemental Nutrition Assistance Program. EBT cards are name based, and the government delivers new balance to the eligible person's card. There is still no particular restriction on who can use the card to make purchases: there is no requirement of checking ID, and since the benefits are given to the household, it is common that a different member of the household picks up the food. Nonetheless, transferring the benefit to someone else became burdensome. One cannot simply hand another person $5 worth of food stamps. If you hand over your card, they will get not only the whole balance, but also future benefits.

Finally, consider **mortgages**. A mortgage can also be thought of as a currency with spending restriction. (Funny, we usually don't think about food stamps and mortgages in the same category.) You get a credit from the bank. But you cannot take this money and go shopping, or go on vacation (although, in the US, under special conditions you can take out a mortgage for such purposes). Generally, you can only spend it on a particular piece of real estate. (Yes, it frees your other money reserves to be spent on shopping or on vacation, but you have always had a freedom how to spend that part of your budget.) Moreover, the mortgage cannot be easily transferred—it is restricted to a single entity (usually a person or a couple). Generally, there are also further restrictions in using a mortgage: since the collateral is often the particular real estate in question, the mortgage holder may be required to provide some insurance for the

property. There may be specific payment schedules imposed and penalties for late and/or early payment. Clearly, a mortgage is a quite complex form of payment.

In all these examples, the special-money system has been introduced by specific institutions (a consortium of shoppers, the government, or a bank, respectively) with specific objectives in mind. The rules governing the currencies create particular incentives for members of the target population. The design features of these private currencies need to take into account these incentives in order to support the organizations' objectives. This often means considering careful trade-offs. For example, in the case of food stamps, the government realized that it needed to restrict transferability to make sure that only the target social group benefits from the subsidies. Too much restriction on transferability, (e.g., providing access only to the head of household) however, makes the use of the subsidy unpractical as it is often family members who are in charge of shopping. The EBT debit card solution is a good compromise in this case. Another important consideration in the choice of features for a currency is the cost of implementing (enforcing) the features. Most of the time, these costs are not trivial (no matter who bears them). Referring to the food stamps example again, the bank fees charged for the management of the EBT cards absorb some of the value of the food stamps, so using EBT cards to restrict transferability has a cost to the currency's issuer (and user).

In what follows, we argue that digital currencies provide much more flexibility in introducing design features and make the monitoring of their corresponding restrictions much less costly. This should mean that the type of businesses capable of introducing them effectively can use them to their advantage, and expand. Nonetheless, the change that digital currencies bring is quantitative rather than qualitative. It is not a completely new thing, but rather a change in degree. Yet, we shouldn't dismiss it because of insufficient novelty. It can still have a very large impact. Email is an example of such quantitative, not qualitative change that nonetheless significantly impacted our work and life in general, creating a more connected and "just-in-time" work place (for better or worse). Email is just quicker mail. Instead of days, we get mail electronically within minutes, or less. In the beginning, people checked their email once in a while, and wrote emails similar to traditional letters. (Remember when you needed to connect via dial-up to collect your email, maybe once every few days?) Quickly, however, messages were more frequently returned and got shorter. Internet connection got better (also because there was

a demand for it). And now we usually send short, informal messages all the time, and receive them almost instantly. Email turned from digital version of letters into digital version of notes passed in a class. Similarly, who knows where the proliferation of differentiated digital currencies may take us?

As the examples above demonstrate, in exploring more recent private electronic money systems we need to keep in mind that such systems are driven by the interplay between the objectives of the organizations introducing them and the incentives that they provide for the users. In turn, in the case of private companies like Internet platforms, organizational objectives are driven by these companies' business models. It is not surprising then that fundamentally different business models lead to private currencies with very different design features.

3.2 Platform-Based Currencies in the Digital Era

The digital era offers unprecedented extent to which the design may be manipulated. In certain cases, technology has also significantly reduced the cost of implementing alternative designs. In particular, technology can easily adjust the three fundamental design features we have reviewed above. It allows, for example, to easily control to whom the currency can or cannot be transferred (i.e., transferability). Technology can also better control how the currency can be acquired (acquirability), and how it can be spent (redeemability). Table 3.1 provides a few possible design combinations that have been implemented by some digital platform businesses. Using the design elements that we have discussed above, it summarizes a few observable combinations of these features in businesses. As can be seen from Table 3.1, these three characteristics definitely seem to differentiate among platform-based digital currencies observed in the real world.

Importantly, each of these three features clearly provides a few specific incentives for their users. For example, if the currency cannot be redeemed for fiat currency, but only be "spent" on the platform, then this reinforces customer captivity or customer loyalty: leaving the platform means leaving assets behind. Surely, it is better to use these assets to consume more on the platform than losing them. This, of course, is beneficial for the platform if its business model is based on usage intensity, for example. This is why many of the digital platforms and

Table 3.1 Design attributes of platform-based currencies

Acquire	*Transfer*	*Cash-out*	
Buy only	No	No	A "wallet" to store cash to be spent on the platform only (Play Station Store Wallet). It often facilitates the administration of a promotion (Amazon Coin)
Buy only	Yes	No	A wallet combined with a trading system, but still only in-platform (Steam gaming platform)
Buy only	No	Yes	A simple wallet that maybe only relevant if there is a need to trigger micro-payments (not observed in platform-based currencies)
Buy only	Yes	Yes	A payment system that does not require a separate currency (e.g., PayPal or Venmo)
Earn	Yes	Yes	A promotion device to encourage people to try the product (not observed in platform-based currencies)
Earn	No	Yes	A job-market (since there is no transfer, there is no need for "buying" (Amazon's Mechanical Turk)
Earn only	No	No	Not really a currency, but may be display of status (DKP in WoW, karma in Guild Wars 2)
Earn	Yes	No	A fully functioning virtual economy with no direct cash-out; one can indirectly cash-out and buy because transfer is possible (Eve Online)
Earn or buy	Yes	Yes	A fully functioning virtual economy (Second Life)
Earn or buy	No	No	A wallet combined with promotion incentives (Facebook Credits)

especially online games (e.g., World of Warcraft) restrict cashing out by simply not making it possible to convert their currencies into fiat currency. At the same time, restricting funds within the platform may also mean that people might be less inclined to inject funds from outside into the platform in the first place (assuming that the platform currency can be purchased with fiat currency). If such "investments" are critical, say for the development of the platform itself, then this consideration needs to be traded off against customer loyalty. This is the case for the virtual world, Second Life, that entirely relies on its users to build all the content on the platform from the texture of the land to plants, houses and any object one can imagine. From the list of three key characteristics, transferability is a particularly subtle one. Transferability is clearly needed if the platform wants economic interaction (trade) between the members.

However, transferability creates possibility for some people to use the platform to earn money and export it from the platform, and this, even if the platform doesn't officially allow such "cashing out." As we will analyze the case of World of Warcraft gold below, it will become apparent that allowing transferability is generally in conflict with strong restrictions on taking funds out of the platform.

What might explain which design features would be implemented for the currency of a particular platform? Based on the discussion above and the few traditional examples that we saw before, it is likely that the platforms' business model will have a decisive role in the choice of features. The platform's business model will provide guidance on the incentives that the platform wants to reinforce for its membership base. Clearly, this may not be the only determinant, there might be many other, maybe practical considerations (e.g., technological or regulatory constraints) but it is safe to assume that the digital currency—if adopted—should support the firm's business model. The concept of "business model" is quite complex, however. To be more specific, we will focus on two of its key aspects: the way the platform creates value to its customers /members (i.e., its value proposition) and the way it captures this value (i.e., its revenue model). We argue that these two aspects of the business model will have a strong influence on the choice of design features for the platform's digital currency.

The dynamic evolution of the Internet has spawned many different business models and the process of experimentation is far from over. At the time of writing this book, we identified four particular models that seem to work for a fairly large number of successful digital platform businesses. These are the following:

i. Online, interactive video games, such as World of Warcraft, Diablo, or Fortnite;
ii. Virtual worlds or metaverses, like Second Life and Eve Online;
iii. Social networks, e.g., Facebook, Instagram, Tencent, and WeChat;
iv. Product promotion platforms, such as Amazon's e-reader platform or a gaming platform like Steam, for instance.

We look at each of these four business models, and analyze their digital currency designs. Our goal is to explore how their value creation process and their revenue model are linked to the kind of currencies that they

introduced. An important caveat is that what we call "typical" business models exhibit a fair amount of variation themselves. In fact, the strict separation of these four categories is somewhat forced as there are many platforms that sit somewhere between the categories. Online interactive video games have an incredible variety from relatively simple and stylized ones to complex universes. World of Warcraft, for example, can be legitimately seen as a virtual world rather than simply a video game depending on one's perspective and we will have to be more explicit in making this difference clear. On the other hand, Eve Online can be seen as a video game rather than a virtual world. Similarly, Tencent can be legitimately seen as a social network even though it is one of the largest gaming platforms in the world, hosting many of its own games. In what follows, we will try to provide a more precise definition for these business models but it is important to keep in mind that any classification is somewhat forced given the large number and variety of digital platforms available.

In one important aspect, these platforms are quite similar; however, they all exhibit some form of consumption externality or network effect. In such environments, consumers benefit from other consumers using the same platform. As the main purpose of platforms is to facilitate transactions between groups of consumers, it is quite natural that such consumption externalities are present. In turn, the recent emergence of such platforms is not surprising given the Internet's core capacity to provide interactivity for large number of people. As such, platforms built on the Internet naturally exploit this feature. Let us take the case of video games, for example. Here, the more people play the game the more enjoyable it is: it results in more thrills and also in more opportunities for collaboration. Similarly, on social networks more people sharing content means there is more content to consume. For the individual member sharing his/her content, there is a larger audience if the platform has a larger set of members. In virtual worlds, more members mean a richer and more complex world with more objects present and more things to do. While there are differences across platforms on how exactly these externalities play out, they are present in one form or another on each naturally leading to network effects (sometimes indirect network effects). Moreover, the presence and nature of consumption externalities is often reflected in the design of the currencies they use.

Finally, it is important to point out, that digital platforms and their currencies are relatively new phenomena, and what we observe today is

not the final and definitive form of the digital currencies of these businesses. The experimentation in this domain is far from over. In fact, and most interestingly, some of the currencies introduced didn't work out (and had to be abandoned) or needed substantial redesign during their short history. These cases are particularly insightful for understanding the link between the digital currencies' design features and their role in the platforms' corresponding business models.

3.2.1 Online Video Games—The Case World of Warcraft Gold

For a long time, World of Warcraft has been one of the most popular massive multiplayer online role-playing game (MMORPG or sometimes just called "morpeg"). Created by Blizzard Entertainment, it still has over 6 million gamers interacting with their avatars in this medieval virtual world. As they play, they gain skills and wealth. They go on quests, alone or more commonly in groups to face challenges and gain more skills and wealth. The quests are demanding and it is important for success to build a team with the right composition of complementary skills for the particular challenge. The currency of the realm is World of Warcraft gold (WoW gold). It can be freely transferred between members of the game. But according to the rules of the game, it cannot be acquired in exchange for fiat currency, nor can it be redeemed for fiat currency. WoW gold can only be earned in the game, and only spent in the game.

It is easy to understand the purpose of most of these design features. Allowing people to earn gold makes them progress in the game. Together with the rule that WoW gold can only be spent in the game, this design creates loyalty. The design is also compatible with the firm's revenue model: a monthly membership fee. The earned and locked-in funds boost loyalty to the game. This is all the more important because the game exhibits strong consumption externalities and associated network effects: the more people play the game the more there are possibilities to form or join teams and complete quests of increased difficulty. In this setup, it makes sense to grow the user base and this is helped by making the platform sticky for those who have already been hooked. Their presence will make the game all the more attractive for new players considering to join.

Transferability is also important for World of Warcraft's value proposition. The game is based on the interactions between players by letting

them form coalitions to complete the quests. Completing a quest is typically rewarded with a bounty. Transferability ensures that the bounty can be appropriately shared across the members of the coalition. This may happen according to skills or contribution to the quest. Transferability also helps members to trade weapons and other objects with one another. This, trading aspect of the game also reinforces the network effects.

Yet, in one respect, this currency design seems overly restrictive: WoW gold cannot be bought with real currency, only earned in the game. Why wouldn't Blizzard want to make extra bucks selling WoW gold? Wouldn't it attract even more members? It turns out that this could actually undermine World of Warcraft's value proposition to members and, as a result, Blizzard's revenues. World of Warcraft's revenue comes from gamers' subscriptions. They keep paying as long as the game delivers the high-quality satisfaction they have signed up for. As mentioned earlier, interaction with other gamers in this virtual world is crucial in the game. The quests at higher levels require several or even few dozens of gamers to collaborate. However, beyond the size of the team, the skills of one's collaborators are also critical for success. Higher level skills are desirable and skills need to be complementary within the team. Yet, most of the time, when selecting team members for a quest the gamer does not know well the potential candidates. Fortunately the status achieved in the game—which can only be guessed from his visible clothing and accessories—is a good proxy for the gamer's skill. A successful quest requires a team with the right mixture of skills. If all status signs are earned by progressing through the game, then status is a good indicator of skill. If, on the other hand, the clothing and accessories were purchased with fiat money, the displayed status no longer correlates with the skill, and is not only uninformative but actually misleading to the gamers trying to put together a successful quest team. If Blizzard were to change the rules, and allow new (therefore unexperienced) gamers buy status from others, this would create a strong negative externality for the other ("honest") gamers. The presence of such impostors can quickly destroy the game if trust in peer players' skills is broken.

WoW gold is purposefully designed to serve the game's business model. It illustrates how deliberately restricting certain attributes of the currency may help creating value to customers. It is important to realize, however, that not all interactive games have these restrictions. In particular, a very large proportion of games (most social games on mobile phones, for example) adopt the classic "freemium" model. In the freemium model,

one can play for free, earning "credit" (typically some digital currency) when advancing in the game, i.e., achieving higher status. Clash of Clans, developed by the game studio Supercell, is a good example. It can be played on a PC or on a smartphone. The game is fairly simple. Players own a village and their goal is to develop it as much as possible. Development essentially means building an army with sophisticated weapons and solid defenses against raiders. Funds to build the army come from economic activities of the village, which—with a bit of oversimplification—boil down to digging for gold. Gold in turn can buy more weapons and so on. An interesting part of the game is that one can use its army also to raid other villages and steal gold from them. In this way, everyone is fighting everyone trying to achieve a better "status" (measured as a rank across players) by adjusting their strategies in terms of the investments they make in their armies, defenses, and gold-digging technologies. However, player can also buy credit with fiat currency that will accelerate their advancement by providing them with extra funds. Instead of a fixed subscription fee, it is these purchases (usually coming from a very small proportion of the players) that represent the core revenue source of the game. Clash of Clans is not really special with this business model. It is a typical freemium game.

Interestingly, freemium games do not seem to suffer from the fact that some players can buy into "status." This is indeed the case for many such games. In these games, while players can interact in various ways (helping each other, trading with each other, etc.) they do not rely so critically on each other's advanced skills. In fact, in Clash of Clans, a partner with a large army is just as good as another no matter whether the army was purchased or "earned" via conquests. As such the value of a partner is not closely linked to experience in the game. In other words, the fact that observed status and skill are little correlated doesn't hurt the other players. Inexperienced players do not create a strong negative externality for others. In the case of Fortnite, probably the most popular freemium game as of writing, such externalities aren't even present. In its Battle Royale game, the so-called V-Bucks game currency can only be used to buy ornamental or esthetic features (essentially, they help the player look "cool" to others) so buying game currency with fiat currency is not a problem: someone looking good does not mean that they are skilled.

Moreover, in Battle Royale, typically everyone is against everybody else so looking for skilled partners is not an issue.[4]

As in the case of morpegs, freemium games typically also have restrictions on the withdrawal of funds created in the game; these funds cannot be retrieved for fiat currency. This keeps players in the game, which also helps other players join. Yet, as we will see below, it is not always easy for the owner of the game to enforce this rule and some players go out of their way to break it.

Challenges: Fraud with digital currency

While games' digital features are easy to implement and monitor on the platform, the Internet has made it increasingly easy to break the rules using other interactive platforms that facilitate commerce, such as e-Bay, for instance. In World of Warcraft, for example, despite rules of the game to the contrary, there are a lot of external transfers between players. People are willing to buy WoW gold for fiat currency (along with items you can buy with gold, like weapons or armor) to advance in the game without the time investment needed. eBay has banned trading of in-game currencies and assets in January 2007,[5] but there are a number of other sites where one can buy WoW gold for fiat currency (e.g., www.goldah.com). In China, DD373.com is an e-commerce platform specialized to trade in-game items. In 2021, it was sued by Tencent, the largest gaming platform, who claimed that it is the sole owner of all digital assets in its games and therefore external trading of these should not be allowed. As of writing, the lawsuit has not concluded but game developers are eagerly following it. Nevertheless, it is clear that there is robust demand for in-game assets. Indeed, there is so much demand that some people in the developing countries turn it into their day job to play games and sell the in-game assets on the "open" market (e.g., collect WoW gold and then sell it for fiat currency). This activity is popularly called "gold mining." In extreme cases it even led to infamous instances of forced gold

[4] Fortnite has a "team play" mode but team members are recruited outside the game (essentially friends enter in the game as a team) or they are randomly matched to the player.

[5] e-Bay probably decided to ban such trading to avoid legal suits. Technically, such in-game assets are property of the game, unless specifically otherwise stated. For example, in Second Life, another "virtual world," the individuals are the owners of their in-game digital assets. Accordingly, trade of Second Life assets is allowed on e-Bay.

mining in Chinese labor camps, where the guards made the prisoners play the game by night, selling the proceeds.[6] Clearly, the existence of these "black markets" did not help the reputation of the game. But even without these extreme cases, as we have seen above, black markets hurt the gamers by flooding the platform with people whose status displays did not match their skill, thereby spoiling the game. Interestingly, World of Warcraft gamers themselves started policing such suspicious behavior and reporting it to the game administrators. As a consequence, Blizzard has expelled several players for fraud but the practice hasn't disappeared.

Gamers themselves found a better solution. The problem of "fake players" became so annoying for them that they decided to ignore the traditional displays of status, and instead were relying on, so-called Dragon Kill Points (DKP) for the assessment of skill of a potential quest mate. DKP are acquired by participating in a quest that kills a particular type of creature (called a *boss*, initially bosses were mainly dragons, hence the name). The killed creature leaves behind a treasure, or a loot. However, when there are many people in the quest, the issue of how to divide the loot is a problem. Games[7] solved this problem by allotting the DKP to the participants of a successful quest, and allowing them to use those points to buy certain items (only those items can be bought with DKP that are rewarded for killing a boss). If a player does not spend their DKP, they accumulate and can be spent later. In other words, DKP are an alternative currency with restrictions on what they can be spent. What is most important, DKP are non-transferable. DKP have far more limited use than WoW gold, and cannot substitute it in its economic role in the game. However, in the presence of back markets for WoW gold, DKP turned out to be more useful for signaling skill. At first DKP were informally assigned and tracked within groups of gamers, called guilds, for the purpose of goods allocation. As they gained importance as a skill signaling tool, Blizzard formalized this dual system as a so-called Guild Advancement system aside of the existing gold.

[6] See www.theguardian.com/world/2011/may/25/china-prisoners-internet-gaming-scam.

[7] DKP originated in the game Everquest in 1999, but were since then also adopted in many other morpegs, including World of Warcraft.

Unsuccessful experiments and learning

Blizzard also experimented with different design of currency system in other games. Not all the experiments were successful. Given the rampant sales of in-game gold and items in the "black market," Blizzard decided to build such functionality directly into the game for the third edition of their popular Diablo game. In this game, a player defeats enemy creatures at increasing challenge levels. At each level a defeated creature drops weapons (and gold) that help the player to defeat a more demanding creature. The culmination of the game is fight with Diablo, the "lord of terror." The game Diablo is less interactive than World of Warcraft, but still has some interactive (cooperative) elements. A player who has a surplus of one type of weapons, armor, or other items, but needs a different type, can trade with other players either directly or through the in-game market (auction house). The trades could happen by using in-game gold, or using real currency. Blizzard charges a transaction fee on such trades (whether they are done with in-game or real-world currencies). Additionally, Blizzard charges a cash-out fee if a player takes out fiat money outside of the platform. Unlike World of Warcraft, Diablo is not subscription based. There is only a one-time fee of purchasing the game. Thus, making money on users' cash transactions in and out of the game made sense from the perspective of the revenue model. However, the possibility for inexperienced players to buy status represented an important problem even in this game where the level of cooperation (and therefore, the assessment of a partner's skill) is not so critical. The endorsement of fiat currency trade within the game eliminated outside black markets and made the company earn more revenues but it also reinforced the negative externalities represented by "fake players." In March 2014, Blizzard closed down the auction houses, saying that the real money auction houses (RMAH) were detrimental to the game because they "short-circuited" the title. Specifically, they said: "the auction houses made the game less satisfying to play as they undermined the challenge of the battle to defeat Diablo, the 'lord of terror'."

Blizzard seemed to get a good grip on the problem of currency design in yet another game called Guild Wars 2. In this game, there are three types of currency: gems, gold, and karma. Gems are directly linked to the fiat money. Players can buy them with fiat money at a fixed rate. Gold can be earned or bought with gems. However, the gold is bought in the player-driven market, the gems–gold exchange rate is not set by

the platform but rather depends on the relative supplies and demands of gold and gems. Karma, on the other hand, is earned through game tasks, cannot be bought or transferred. Most micro-transactions and in-game purchases, whether from the platform or directly between players, occur in fully transferable gold. But karma is used to buy unique awards. Thus, while gold (and indirectly gems) can be used to adorn avatars and the players' environment, only rewards bought with non-transferable karma directly signal the player's skill.

These examples of Blizzard games show how important is the impact of transferability for the currency. Transferability makes it easier to bypass other restrictions, like restrictions on buy-in and cash-out, if users find it beneficial to do so. If it is important for the value of the platform that the users do not buy-in the currency, like in the case of skill signaling in World of Warcraft and Guild Wars 2, the platform needs to rely on currency that is not transferable. It may create conflict if at the same time, one of the attractions of the activity on the platform are economic interactions between users. These economic activities on the platform usually require a transferable currency. As Blizzard's example shows, one possible solution is a dual system that allows for buying in the transferable currency, formalizing a de facto state, while also operating a separate non-transferable currency earned in the game and thus signaling the skill.[8]

3.2.2 Virtual Worlds and Linden Dollars

The previous examples looked at restricting currency functions. But the optimal design for some platform businesses may point to a fully equipped currency. An example of such an unrestricted currency is the Linden Dollar, the currency used in a virtual world called Second Life. At this point, it is important to ask: what is the difference between a "virtual world" and a complex video game such as World of Warcraft? The short answer is that virtual worlds are "MMORPGs without purpose." Classic morpegs represent a well-defined world with well-defined rules and a consistent visual appearance. Most importantly, they have well-defined

[8] It is still possible to transfer even currency that was designed to be non-transferable, if the users have strong incentive to do so. One way is to transfer the entire character, which requires sharing login and password. But it is much less convenient for both the buyers and the sellers.

goals for their players. Players face specific quests, there is a known hierarchy among them and everyone knows what needs to be done to achieve the goals. In Second Life, pretty much nothing is defined. One can choose to do whatever she wants and people end up doing wildly different things. These can be very complex activities such as running a virtual bar (yes, with virtual drinks and real music mixed by a DJ represented, of course by his/her avatar), building and selling sophisticated spaceships or operating a gallery with beautiful paintings. In contrast, activities can also be really simple and mindless such as hanging out with friends (maybe in a bar), decorating one's avatar, or just visiting locations in the virtual world, etc. In fact, the virtual world itself is pretty undefined too—one can visit entirely different universes and meet avatars with totally differing looks. In one region, for instance, some members have rebuilt the entire universe of the movie Avatar with floating islands and spectacular vegetation. Other regions were built to look like abandoned industrial wastelands. Every aspect of the environment, from the shape of the land, to the vegetation, buildings, creatures, etc., need to be built from scratch by the members of the virtual world.[9]

Second Life is probably the most extreme of virtual worlds in that almost nothing is defined in it—it provides unlimited possibilities. In this sense, it is the opposite of World of Warcraft that is a fully codified virtual world. Virtual worlds represent a continuum between these extremes with many platforms sitting somewhere in the middle. Eve Online, for example, is a utopian virtual world inspired by Ayn Rand's ideas of a libertarian universe. While it also allows almost unlimited freedom to its members, the environment is somewhat more defined than Second Life's. In one respect, however, Eve Online is "freer" than Second Life: it has no property rights enforced by the platform. Instead, members need to get organized to enforce these. In Second Life, property rights are well policed by the game and only by hacking the platform can someone steal virtual property form others.

Given this freedom, there will be many different people on the platform with wildly differing tastes and very different activities. In fact,

[9] The platform provides very sophisticated building and scripting tools to allow the creation of extremely complex objects which may interact with each other and with the avatars in a sophisticated way. One could speculate that part of the challenge for Second Life in attracting a large number of users is the complexity of this user interface. The creation of complex objects requires very special skills and hours of work.

virtual worlds are built to be full blown economies. This also means that these platforms have many opportunities to collect revenues from their members. A general tax on economic activity seems to be a good way to collect revenues. In the case of Second Life, where the world is virtually empty without the objects built by residents, a good proxy for economic activity is land ownership (which only makes sense if something is built on the land). Taxing virtual land is also compatible with the platform's cost of serving customers (more land and more objects on the land mean more memory used by the platform's IT system).

It is not surprising then that Second Life's revenue model is tightly linked to the total economic activity of its c.a. 1 million regular users on the platform. According to estimates this "GDP" amounts to about half a billion US dollars. Specifically, Linden Labs, the owner of Second Life, has three main revenue sources. First, it collects revenues from "advanced" users (basically, those who can own land and can build things). Essentially, it charges a membership fee (of about $12/month in 2021). Second, Second Life also collects revenues with the sale of virtual land. Depending on the size and properties of the land in question this may cost a few hundred dollars with a monthly fee (usually around $200). Finally, Second Life also collects a transaction fee from the exchange between Linden dollars and fiat currency. All of these revenues are broadly linked to the diverse economic activities on the platform.

Second Life's currency, the Linden dollar, is a fully equipped currency (in May 2021, for example, 1 US dollar traded for a little more than 300 Linden dollars). It can be earned within the platform, usually by working for someone, but it can also be purchased with fiat currency. It can be transferred to anyone or spent within the platform to purchase anything that is for sale. Finally, it can also be changed back to fiat currency and taken out of the platform. This last feature is somewhat puzzling. Why does Second Life allow people to take out their money from the platform? As we have seen before, this may encourage people to leave the platform either because, simply, there is no cost in leaving when one wants to try other things or because people may want to "cash out" after having been successful. Clearly, this does not necessarily benefit Second Life, especially in view of the strong positive consumption externalities present for virtual worlds. Why allow it then? In short, the "cash-out" policy has to do with the provision of incentives to "invest" in the platform by building content in it. If Linden Labs wants to have a vibrant interactive community on the platform, including a complex economy, it needs to provide incentives

for people to invest and this for a very heterogeneous membership base. First, people need to build things. For complex objects (e.g., a plane, a musical instrument, or a shopping center) this might require the collaboration of multiple people or the combination of multiple elements already available from others. Since a common (set of) "quest(s)" are not available, collaboration often necessitates the hiring of labor. Furthermore, just as in the real world, to function well in Second Life extensive trade is required. Most complex objects also require large investments of time. It is unreasonable to expect, however, that everyone can spend this time in the game, so the platform needs to encourage investment in terms of money. For all practical purposes, Second Life is like a real economy, with investment, labor, and product markets and clear property rights. Indeed, people pointed out that one of the reasons why users were willing to build a large variety of things to populate the virtual world was that, very early on, the platform declared that the residents (as Second Life users are called) owned the virtual assets created in the game and they could freely sell these.

In this sense, Second Life is not a game like the morpegs we have reviewed above. By 2008, many of its residents moved (part of) their professional lives to Second Life earning Linden dollars by building things, opening and running stores, or simply working for other virtual businesses. Accordingly, the IRS declared that earnings in Linden dollars were taxable and many other governments made similar announcements. While for most people these earnings were ephemeral, some people made a real fortune selling digital goods on Second Life.[10] Many real-world businesses (from retailers like American Apparel, media companies such as Reuters to, maybe more naturally, technology companies like Sun Microsystems) decided to build a presence on Second Life. They were followed by other organizations (schools, universities local or national governments) starting serious activities[11] in the virtual world with the hope that it will eventually become a dominant Internet platform. While these hopes have since largely evaporated as Second Life remained a

[10] One of the famous characters of SL, also called the "First Virtual Millionaire," was Anshe Chung who established a successful operation, developing, renting, and trading virtual real estate. She was reported to have earned over $1 million.

[11] For example, American Apparel opened a showroom, Reuters hired an in-world reporter and the Swedish embassy has opened an office in Second Life.

relatively small platform compared to, say dominant social networks, its current c.a. 1 million residents still sustain a vibrant economy and an extremely diverse virtual universe.

Eve Online

The virtual world, Eve Online, mentioned earlier, is also a quite typical virtual world in that little is defined for its members who can freely choose their activities. While there are a few important differences (e.g., the whole setting, which in Eve Online is a science fiction space setting) the platform's members' activities add up to a fairly complex economy that, in many respects, is even more free than Second Life's. As mentioned earlier, property rights are not enforced centrally and, instead, members need to get organized to protect their properties by hiring fighters, etc. Eve Online is more "market driven" than Second Life in the sense that trade—as opposed to user-generated content—constitutes a more important part of the game. While there are more constraints on creating something in Eve Online, trade is more complex and requires special skills acquired in the game. Essentially, it boils down to the ability from the trader to see more arbitrage opportunities than other players who may have invested in developing other skills (fighting, building, etc.). All these small differences, however, do not really matter for the big picture, namely that both platforms run a full-fledged, complex economy. Consequently, both virtual worlds need to provide investment incentives for their members. Accordingly, it is not surprising that Eve Online also has a fully equipped currency. It is denominated in ISK, which is somewhat confusing, not just because the Icelandic krona is also abbreviated as ISK but also because Eve Online's developing company, CCP Games, is based in Reykjavik.

Impact on the real world

While both Second Life and Eve Online have introduced fully equipped currencies—there are no restrictions on buying, earning, gifting, or transferring them or even changing them back to fiat currency—neither Linden dollars nor ISK have had significant impact outside of their respective platforms. Probably, the main reason is that neither platform managed to attract a very large community. There are millions of Second Life accounts registered, but as of writing, it is estimated that around a million residents represent active players. Eve Online's active gaming population is

estimated to be about 300 thousand players. Clearly, these numbers are dwarfed by the billions of members on Facebook, for instance.

Yet, early commentators were mostly worried about the fact that virtual worlds' currencies were fully equipped, thereby having the potential to replace national currencies. However, it is misguided to be concerned about the impact of the currency outside of the intended platform just because it is fully equipped. As Fung and Halaburda (2014) show, the currency does not need to be fully equipped to have a potential for impact outside of the platform. It only needs to be transferable. Once transferable, the restrictions around acquirability and redeemability can be manipulated by the users. This was the case for WoW gold, which was broadly traded outside of the World of Warcraft platform despite lacking some key features that Linden dollars had.

With full transferability, if people want to acquire or redeem the currency for fiat currency, they can find a way for mixed trade, where one part of the transaction takes place on the platform, and the other outside of it. Once the currency is traded outside of the platform, it could be used for trades other than intended by the platform. Whether it will have an impact outside of the platform boils down to whether users have an incentive to use it instead of the currencies already available. In such a case, it becomes a typical (pre-digital) issue of currency competition. It is like competition between US and Canadian dollars. Even though there are no restrictions, Canadians have no need to use US dollars in Canada, or Americans to use Canadian dollars in the USA. But there may be some countries (Argentina is often pointed out as one of them), where people prefer to use US dollars instead of the local currency (Peso). The same dynamic plays out with digital currencies. Even with transferability, they only become adopted outside of the platform if they serve some functions better than existing alternatives. As far as we know, it hardly happens for WoW gold, or for Linden dollars. But there is one, well recorded example where it did take place: Q-coins, the currency of Tencent, a Chinese social network, which we present later in this chapter.

3.2.3 Social Networks and Facebook Credits

Social networks are the third prototypical business model that has emerged for Internet platforms. On these large platforms with hundreds of millions of users, members interact mostly by sharing content with

one another. The revenue model is usually advertising based, although there have been other sources of revenues also providing significant contributions (e.g., revenues from app developers or game developers).[12] Facebook is by far the largest social network in the world with over 2 billion active users. It also owns a variety of other leading social platforms that are more or less connected to Facebook (e.g., Instagram, WhatsApp, or Facebook Messenger). It is important to realize that Facebook is not simply a platform for its members to interact with user-generated content. It is a so-called multi-sided platform, where a lot of the content is provided by third-parties, be it media sites, game or app developers, news organizations, or simply product brands. Popular categories of this, third party content consists of videos, articles, games, etc.

In 2009, Facebook introduced Facebook Credits, which became the mandatory currency in 2011 for all apps and games on the Facebook platform that wanted to charge members. Facebook Credit used non-USD denominations and essentially functioned as a virtual wallet. You could add funds online or by purchasing gift cards at big box stores. The system has since been retired in 2013 in favor of a payment system using USD.

As mentioned earlier, Facebook Credits could not be transferred between Facebook users. They also could not be exchanged into fiat currency, like dollars, Euros, or yens.[13] They could be spent on anything on Facebook, whether the content was directly provided by Facebook or by a third party developer, as long as developers accepted Facebook Credits. Between 2009 and 2011 developers could charge in Facebook Credits or fiat currencies. From 2011 until 2013 the developers no longer had a choice, and had to use Facebook Credits if they wanted to charge the users.

In terms of acquirability, users could buy Facebook Credits using fiat currencies (the price was about 10 Facebook Credits per dollar, with a number of quantity discounts; e.g., for $10 there was 5% bonus, and one received 105 Facebook Credits). Interestingly, users could also *earn* Facebook Credits, for example, by testing a game or taking a survey. Gans and Halaburda (2015) show how restricting the currency's functionality

[12] In 2020, well over 90% of Facebook's $86 billion global revenue came from advertising. The remainder came from "payments and other fees."

[13] Facebook calls them regional currencies, trying not to get into whether they are national or state currencies, which could be imprecise, e.g., in the context of the Euro.

in such a way was optimal for Facebook. The answer is related to the fact that Facebook's main source of revenue is advertising. The advertising revenue is directly related to the time users spend on the platform. And Facebook Credits were optimally designed to induce users to spend more time on the platform.

An important driving force comes from the fact that "consumption" of social network content exhibits consumption complementarities. That is the more time my friends spend on Facebook or Instagram, uploading pictures and videos, writing posts and commenting on my photos, the more fun for me it is to spend time on these social networks, either posting my own content or commenting on my friends' posts. This in turn gives rise to positive network effects—the more people are active on the social network the more utility I am getting from spending time on it. As we saw, it is a very common property for Internet companies, and a very valuable one. If Facebook can induce one user to spend more time on the platform, it has a multiplier effect due to these consumption complementarities, as it will induce other people to spend more time, and maybe new people to join Facebook.

Facebook Credits were designed to induce users to spend more time on the platform, which further induced other people to spend more time. Facebook Credits gave users a way to enhance their Facebook experience. For example, with Facebook Credits users could send virtual flowers to a friend, could gain additional options in a game, e.g., fertilizer for virtual plants to increase the "harvest" in their virtual farm, or feed for their virtual pet. All those activities made spending time on Facebook more pleasurable, and thus induced people to spend more time. By allowing both buying and earning, Facebook made sure that Facebook Credits were accessible both to users who had more money than time on their hands (cash-rich), and those who had more time than money (time-rich).

In turn, allowing for transfers between users or exchanging Facebook Credits for fiat currency could only undermine this objective. Allowing for exchanging Facebook Credits for fiat currency would allow users to sell earned Facebook Credits back to Facebook. Allowing for transfers between the users could result in a situation where time-rich users earn and sell Facebook Credits to cash-rich users. To make sure that cash-rich users would prefer to buy their Facebook Credits from them rather than Facebook directly, time-rich users could charge lower price than Facebook's official rate. In both of those cases the time-rich users could earn

Facebook Credits, and sell them instead of using them on increasing Facebook activity. It is true that users spend time on Facebook while earning Facebook Credits. But this activity mostly does not contribute to advertising revenue. And on top of that, if time-rich users do not spend more time on Facebook activities, Facebook is losing the multiplier effect of attracting other users to spend more time. Thus, equipping Facebook Credits with these other attributes would be less beneficial for Facebook.

Why were Facebook Credits shut down?

Interestingly, Facebook Credits were phased out at the end of 2013. Was it because they were badly designed? From Facebook's perspective, not necessarily. For sure, at their introduction, users complained at the added level of complexity. Many of the Facebook apps already had their own currencies. For example, Zynga, a large game developer had zCoin as its internal currency that could be used across Zynga games. After Facebook Credits were made mandatory for apps to use, users needed to exchange their dollars into Facebook Credits, and then Facebook Credits into zCoins or FarmVille Dollars. Facebook tried to push the app developers to use Facebook Credits as in-app currency, but with no, or very limited success. App developers, like Zynga, preferred their own currencies, because this locked users to their particular app. Facebook Credits, conversely, could be moved between apps. Many of these apps were games that, as we saw above, cared a great deal about consumer loyalty. In other words, by requiring all apps to use Facebook Credits, Facebook tried to make consumer switching between Facebook apps easier. In this way, Facebook created increased competition for its app developers.[14]

More competition between app developers could have been a good thing for Facebook members. With lower switching costs, this could have encouraged users to consume more content on Facebook, which in turn could have led to even more advertising revenue, etc. However, this argument doesn't take into account that Facebook is a multi-sided platform. To create a healthy ecosystem of apps it needs to provide sufficient incentives for app developers to invest in quality content. If too little surplus can be captured by app developers, these may seek revenues elsewhere,

[14] It is interesting that Tencent, whose revenue model relies on charging users for social games (it is the largest gaming platform in the world), uses a digital currency exactly for this purpose: to make sure that users remain loyal to Tencent not necessarily to individual games within its platform.

leaving Facebook and, overall, contributing less content to the platform. Zynga, for example, has been one of the largest developers of Facebook games and, early on before advertising was really turned on, it was the largest source of Facebook's (initially small) revenue. However, Zynga as well as most other game developers had their own platforms operated outside of Facebook. Interestingly, the same reasoning that made Facebook Credits beneficial for Facebook, prevented app developers to switch from their own currencies to Facebook Credits.

Tencent's Q-coin

Tencent is a Chinese social network that is quite different from Facebook. Its revenues are mostly coming from the sales of digital goods that people use to build avatars, to decorate their page, play games, or to give digital gifts to each other. Tencent does little advertising and, in this particular respect, it is very different from Facebook. While it introduced a currency, called Q-coin, in fact, earlier than Facebook a notable difference from Facebook is that members can use Q-coins outside the platform. While this wasn't a feature that Tencent introduced on purpose, the company supported it when it realised that Q-coins are used outside its platform.

Although it is a social network, in many respects Tencent resembles a freemium game, where players can participate for free but can "buy" a better experience if they spend money on the platform. When someone opens a profile he/she starts earning Q-coins in proportion of his/her activities. These Q-coins are provided by the platform itself. She also earns a status that is linked to her "influence," which depends on the appearance of her page, her activities, and connectedness. People can also buy Q-coins with real currency—essentially buying status, which is again similar to freemium games. Indeed, only a small proportion of the members buy Q-coins, yet these are responsible for a significant part of Tencent's revenues. People can also use their Q-coins to play a large array of games on the platform. Indeed, Tencent is one of the largest gaming platforms in the world, itself developing many original games.

Tencent's Q-coin was introduced in the early 2000's. Q-coins can be earned or bought and they can also be transferred between the users. But they cannot be—officially—exchanged back for state currency. Despite this, unlike other examples of digital currencies, Q-coin gained traction outside of its own platform, creating a fair amount of controversy between China's central bankers, although until 2009, no formal action was taken

against Q-coin.[15] Originally, only intended for purchases of virtual goods and services, like electronic greeting cards, cartoon portraits, chips in online QQ games and anti-virus software, it became popular for peer-to-peer payments. At the beginning, people used it between close friends for simple transactions, like splitting a bill in a restaurant or for sending cash gifts to each other (a popular Chinese custom). Gradually, online merchants started to accept Q-coins as payment. Some brick-and-mortar merchants followed as well. It was reported that you could buy groceries or get a haircut and pay with Q-coins through your Tencent account (Fowler and Qin, 2007).

As the value of the trade using Q-coins was increasing, People's Bank of China (Chinese central bank) expressed concerns about Q-coin's impact on the yuan (since 2006). Tencent managers pointed to the restrictions on the currency's functionality as important mitigating factors. In February 2007, the Shanghai Daily reported Song Yang, an assistant public relations manager at Tencent saying: "The fact that the Q-coins cannot be officially changed back into money makes them less than harmful to the financial market." However, as we mentioned before, full functionality is not necessary for a digital currency to have impact outside of its intended platform. Instead, a necessary (although not a sufficient) condition is transferability. With transferability, users can indirectly redeem Q-coins by transferring them among themselves inside the platform and exchanging national currency outside the platform. As we saw, this was the case for the black markets for WoW gold. Furthermore, if the Q-coins are redeemable for goods and services, there may even be no need to exchange them for national currency.

Indeed, the trade using Q-coin continued to increase. The value of trade in Q-coin reportedly reached several billion Renminbi by 2008. The following year, the Chinese government introduced regulation banning the exchange of a digital currency for real goods and services, in order to "limit its possible impact on the real financial system." Almost two decades after its introduction, there is still a secondary market for Q-coins where people sell them for real money. Q-coins can also be transferred to others. On secondary markets people accumulate larger sums and sell them back at a slightly advantageous rate. At the time of writing this book, on Taobao, we saw a seller offering 50 Q coins for 47.44 rnb whereas the

[15] See http://en.people.cn/200701/12/eng20070112_340681.html.

"official rate" is 1 to 1. But Q-coins aren't that relevant today—in fact, most gaming companies simply charge the official currency and people use their credit cards to pay.

In retrospect, it is clear that the early expansion of Q-coin outside Tencent's platform was fuelled by Q-coins filling an important gap acting as an easy-to-use payment system, essentially replacing credit cards. At the time, when most Chinese people did not have credit cards, e-commerce sites accepting Q-coins made trade significantly easier.

Even for Tencent, Q-coins are of marginal importance today. The history of its transition to smartphones nicely illustrates this. Tencent has developed two approaches to conquer the smartphone. On the one hand, it introduced a mobile app with a version of the well-known existing Tencent platform, adapted to smartphones. In parallel, Tencent also introduced an entirely new social network built from scratch and optimized for the mobile device: WeChat. WeChat is one of the most successful social networks today. Interestingly, Q-coins are not promoted—in fact not even usable on WeChat. WeChat uses fiat currency and functions similarly to Paypal. In fact, in many ways, WeChat is an all-around e-commerce site (a sort of super-app) where all kind of Internet businesses can promote and directly sell their products and services to users (similarly to Amazon, Alibaba, Uber, etc.).

3.2.4 Promotion Platforms and Amazon Coin

The last business model we analyze is promotion platforms. Promotion platforms are specialized two-sided platforms that bring together buyers and sellers. The role of the platform is to facilitate transactions between these groups of customers without really getting involved. Promotion platforms are somewhere between stores and full-blown markets like the NYSE or Amazon's e-commerce platform that provide trading opportunities for a large and diverse set of stocks or products. Rather, they are markets with products that are closely linked. Game platforms hosting a multitude of, mostly similar games (e.g., Valve's Steam gaming platform), Amazon's e-reader ecosystem and Apple's app store are good examples of promotion platforms. Many times, these platforms offer a proprietary currency to their users. One could think of these currency services as virtual wallets. The virtual currency can be purchased with fiat currency but typically, it cannot be changed back to fiat currency. Typically, it is

not transferable (even though one may be able to buy gifts for another person). In almost all cases, this relatively closed system serves some form of promotion activity.

To see this, let's consider a particular example: Amazon Coins. Customers get Amazon Coins when they buy Amazon's Kindle Fire tablet. Otherwise, customers can only obtain Amazon Coins by purchasing them. Amazon Coins cannot be earned. They also cannot be transferred between the customers. This last feature can sometimes create problems, when, for example, Kindle Fire is purchased as a gift. Amazon Coins that come with the tablet cannot be later transferred to the recipient of the gift. The customers cannot exchange Amazon Coins for fiat currency. And they can be spent only on a very limited selection of goods. It is often said that Amazon is the retailer with the largest selection on Earth. But Amazon Coins can only be spent on selected apps on Kindle Fire. And not even any app. The apps need to satisfy some conditions to qualify. They need to take advantage of unique properties of Kindle Fire (as opposed to other tablets running on Android, for example).

Those properties are too restrictive for Amazon Coin to gain ground as a widely accepted currency. Why would Amazon not take advantage of its very large customer base and product selection by introducing a currency internationally—instead of restricting it so much? The answer is that the currency serves a particular promotional purpose. Amazon was a relative late comer to the tablets market (because the non-tablet original Kindle was such a success). The market for tablets is another market characterized by network effects. However, this type of network effect is somewhat different from the ones present among Facebook users. It is more like the network effect between Facebook users and Facebook app developers. We call these network effects indirect. The more applications are available for a particular kind of a tablet, the more valuable it is for the consumers (this is if the quality of the apps is at least the same as there is more of them). But the developers want to develop applications for the kind of a tablet that has the most consumers, as then they will have a larger base to whom the app could be sold. Thus, more apps attract more consumers, which attract more apps, which attract more consumers. And indirectly, the more popular the tablet is—that is, the more consumers have purchased it—the more attractive this tablet to me as a consumer, as it offers more apps. Hence we call them indirect network effects.

It is easy to see how such increasing returns dynamics (i.e., "large grows larger") give rise to so-called "winner-take-all" outcomes. These

forces make it very hard to enter these markets, much harder than entering markets without network effects. Usually successful entry into market with network effects involves subsidizing or "bribing" one of the sides, or initial group of consumers, to gain needed critical mass.[16] In the case of Amazon, lowering the price too much would hurt Amazon's revenue from this category. Instead, Amazon wanted to increase the value of the Kindle Fire tablet by having more apps available for users. However, just having more apps for Android would make all Android-based tablets more valuable. Instead, Amazon needed to procure apps that would be specific to Kindle Fire. One way would be to just pay developers for developing such apps. But that may be risky. How would you know they develop really good apps that consumers will value (after all, they already got their money)? A solution is to give them the money only after consumers "voted with their feet," i.e., purchased those apps. In this way, the most valuable Kindle-Fire-specific apps earn most money. This would give developers incentive to develop better apps. And this is what Amazon did.

Customers who bought the second-generation Kindle Fire for $199, got $50 in Amazon Coins. It may be seen as a rebate, but since the Coins can be spent only on the approved apps, so it is not really money back. They cannot spend it freely. It could count as "money bay" only for customers who wanted to spend $50 on Kindle Fire apps anyway, which is probably not the case for most customers. The developers know that for this particular platform users have $50 that they can only spend on the apps. They will be more willing to spend Amazon Coin than regular cash, so they will be more likely to purchase approved apps. For the app to be approved by Amazon to be legitimate for Amazon Coin, the app needs to demonstrate that it takes advantage of features specific to Kindle Fire (and thus it increases the value of Kindle Fire more than other Android tablets). Just getting the app approved does not guarantee that the developers will get Amazon Coins. Those apps are subject to ratings and reviews as much as any other apps. So consumers will choose to purchase the most valuable of the approved apps. Amazon Coins that the developers collected can be redeemed from Amazon (after the typical cut of 30%). Even though the developers can redeem Amazon Coins, the currency is still non-redeemable for the customers. Thus, Amazon is not giving the

[16] This has been shown in economics research in early 2000s, e.g., Rochet and Tirole (2003), Caillaud and Jullien (2001, 2003).

$50 to the customer purchasing Kindle Fire, but to the developers who make Kindle Fire more valuable.

Easing any of the restrictions would be at odds with this goal. Allowing consumers to exchange Amazon Coins for fiat currency would take away the incentive for the developers, because most people would take the cash, or spend the Coins on other items on Amazon.com that they wanted to purchase anyway. Accepting Amazon Coin in other areas of Amazon business would have the same effect. It would not help start the network effects going for Kindle Fire. And how about transferability? If Amazon Coins could be transferred between customers, it could result in skewed distribution of Amazon Coins. Those few people who use large number of apps would get Amazon Coins from people who would rather spend this currency on other products. Those intensive app users would have lots of Amazon Coins, while most people would hold on to few or none of the Amazon Coins gifted to them by Amazon. But intensive app users would not buy multiple copies of the same app. This would result in a larger number of distinctive apps purchased with Amazon Coin, but fewer copies per app. The best apps would see their market shrinking, while some not so great apps would still be bought by the intensive app users. This would make the whole scheme less attractive for the (good) developers. And, most importantly, it would not provide such strong incentives to produce the best apps. Thus, again, we see that Amazon Coins is a currency optimally designed for the purpose it is supposed to serve.

Steam Wallet dollars

Amazon's transaction platform is specialized in a product (content) category and is not the only such promotion platform, of course. Another category is video game platforms. These platforms offer a collection of video games bringing together players and game developers. For players they provide a convenient store front with search capabilities and a digital wallet that may help them allocate their rewards and funds across various games. For the game developers, they provide an advertising and promotion platform and an opportunity to build loyalty with their customers. As with Amazon coin, it makes sense for the platform to offer a currency that can be spent across games. While, here, people might be able to earn the currency in a particular set of games, the idea is to keep people in the ecosystem for them to spend whatever they have on other games present on the platform.

Steam is an example of such a platform. Originally, it was developed by Valve, an online game developer to provide a site from which gamers could download the updated versions of previously released games. It quickly became clear that the site can also serve as a distribution platform for new games. And once Valve's gamers became regular visitors on Steam for their updates and purchases, the company realized that it could open the platform for third-party developers to allow them to sell and update their games. As an early mover Steam benefitted from indirect network effects: gamers liked Steam because it had the largest variety of games and, similarly, game developers were attracted to Steam because of all the gamers visiting it. By the early 2000s Steam has become one of the leading online distribution platforms. Yet, no matter how large, such a distribution platform does not necessitate a virtual currency. What then lead to the introduction of Steam Wallet?

The answer is user-generated content (UGC) in games. In many online games, users can create small modifications, digital equipment, or new rules within the game that may be shared with other users. Sims, for example, is one such game that gained much more popularity when such sharing has become possible. From a technical perspective, the Steam platform was already well suited to serve such sharing across users. Yet, it could do even better by providing an incentive for the development of UGC via the facilitation of trade. Along the way, Steam can make additional revenues. How does it work? Steam introduced a wallet that users can feed with their credit cards (or offline purchases of gift cards at game shops). Players can search for and acquire a very large variety of UGCs across the games available on the platform. Popular modifications are rewarded by paying the creator of UGC, i.e., by crediting his/her Steam wallet. Steam, of course, keeps a portion of the transaction as revenue.

Steam Wallet dollars are not redeemable for real currency. As with Amazon Coin they need to be spent within the steam ecosystem. As this ecosystem grows very rapidly, Steam has a strong incentive to keep its members spending money on the platform.

3.3 Conclusion

The examples above illustrate how different attributes of currencies induce different usage and behavior of users. And therefore, the optimal set of attributes depends on the platform's business model.

A general design feature of platform-based currencies is not to allow for cash-out. This is directly related to the platforms' effort to increase loyalty and lock-in for their members. This is particularly important for platform businesses because there are strong consumption externalities leading to network effects. A member that keeps spending time on the platform will make the platform all the more attractive to others. This largely explains why most platform-based currencies have no cash-out options. A notable exception is the category of virtual worlds. As we have seen, in this case, providing strong incentives for people to invest in the platform content requires the possibility for members to recoup their cash.

Strong network effects also favor the idea that users can buy-in the platform currency with fiat currency. Again, this can only grow total activity on the platform and in the presence of consumption externalities make the platform more attractive to existing as well as new users. Yet, this argument has a limit in one particular case: when some platform-specific meritocracy is a key part of the platform's value proposition. Then allowing for buying in may disturb this meritocracy and have a negative externality on the users. Indeed, if some sort of in-game meritocracy is important for the functioning of the platform then the platform should refrain from allowing buy-in for their members. This was most visible for the case of morpegs, where skill was important for all players to enjoy the game and one could fake skills by purchasing certain items. In fact, one of the key insights from the past two decades of experimentation is that in-game meritocracy does need to be shielded away from economic exchange. We have seen how this problem can be solved if meritocracy is "shielded away" from the currency used for the in-game commerce, for example, by a dual currency systems.

Transferability is probably the most critical design feature of a currency and the most nuanced. In practice, it is the only feature that is necessary (but not sufficient) for the currency to have an impact outside of the platform. Transferability is necessary if the platform needs to promote economic activities (trade, in particular) for its value proposition, which

is the case for many of the interactive business models. Yet, once transferability is allowed, it opens a back door for users to buy-in and cash-out even if the platform's policy aims to avoid it.

3.4 Are Restricted Currencies Really Currencies?

Most platform-based digital currencies are restricted in at least some of their attributes. Some, like Facebook Credits and Amazon Coins are even restricted in transferability, arguably the most important attribute of a currency. One can legitimately ask whether restricted currencies are still money?

In the previous chapter we have discussed the economic definition of money and its limitations. Money is defined as (1) unit of account, (2) store of value, and (3) means of exchange. So is Facebook Credits, Amazon Coin, or WoW gold money? Some scholars argue they are not. Maybe they have their own unit of account (but are pegged to a fiat currencies), but they are poor store of value and one can hardly use them as widely accepted means of exchange in transactions.

However, WoW gold is definitely a currency—it is the currency of the World of Warcraft realm. You can't use US dollars in World of Warcraft. As such, the US dollar is not a currency in World of Warcraft. The issue with Facebook Credits and Amazon Coin is more complicated. Their transfer, and therefore role as means of exchange is limited. Facebook Credits could only be paid to Facebook. But then again, one could purchase those items only with Facebook Credits, so Facebook Credit was a currency (means of exchange) for this particular transaction. Currency should facilitate trade. It may have broader or more specific applications. But so does the US dollar and the Swedish krona. Currency may facilitate trade in a specific geographical area, or only for a specific kind of trade. The more limited the trade it can facilitate, the more limited the currency. In fact, we can say that especially with the design possibilities opened with digital currencies, we have a whole spectrum of semi-limited and limited currencies.

CHAPTER 4

Bitcoin and Arrival of Cryptocurrencies

So far in this book, we saw digital currencies issued by digital platforms. These innovations deserve the title of "digital currencies," and they are a good object to start explaining the economics of digital currencies. However, when people hear "digital currency," their first thoughts will likely be "cryptocurrencies" and "Bitcoin." This is not surprising: for the past few years these terms have appeared frequently in popular media, technical discussions, and even in policy debates and legislation. We now move on to this second type of digital currencies, flash out the main differences with platform-issued digital currencies, and discuss what implications such differences have for the economics of cryptocurrencies and for their potential widespread adoption.

Before we do this, however, we would be remiss not to discuss the catalyst of the ongoing media commotion: Bitcoin. Bitcoin is a decentralized digital currency invented in 2008 by somebody hiding behind the pseudonym of Satoshi Nakamoto. Nakamoto proposed Bitcoin to address an economic problem inherent in electronic commerce: the frictions and the high transaction costs of trading over the Internet, particularly relevant for small-value transactions. Indeed, while the key innovation in Nakamoto's paper is cryptography and computer science, those who read it often comment on how much space it devotes to economics and the theory of money of the sort we discussed in the earlier chapters of this book.

H. Halaburda et al., *Beyond Bitcoin*,
https://doi.org/10.1007/978-3-030-88931-9_4

In its early years, Bitcoin has been known to a relatively narrow community of cryptography enthusiasts. The first time the currency made it into the mainstream media was probably in June 2011, during the WikiLeaks affair. WikiLeaks is a website that publishes information, especially news leaks and secret information from classified sources. In 2010 WikiLeaks published a number of classified documents related to the war in Afghanistan, which brought mainstream media attention to the site, and put WikiLeaks at odds with the US government. In December 2010 a number of banks and payment services providers (e.g., Bank of America, PayPal, Visa) refused to provide WikiLeaks with their services, making it difficult if not impossible for the website to receive donations from its supporters. WikiLeaks' founder, Julian Assange, decided in June 2011 to start accepting donations in Bitcoin, highlighting the flexibility of the currency, its anonymity, and independence from traditional financial providers.

Bitcoin grabbed the headlines again, in an even more spectacular manner, in late 2013, when it appeared to be an interesting speculative investment opportunity. Its price (i.e., its exchange rate to the US dollar) skyrocketed from below $15 at the beginning of 2013 to over $1,200 at the end of November 2013. At the same time, Bitcoin was gaining a foothold in electronic commerce. For example, Baidu, a Chinese search engine, decided in October 2013 to start accepting bitcoins for Jiasule, its commercial service for improving the security and performance of websites.

On the other hand, another big reason for Bitcoin's presence in the media was its notoriety. The currency was at the center of a number of events and scandals. The biggest of them was the Silk Road raid by the FBI. Silk Road was a website that matched buyers and sellers of illegal substances and services, for example, drugs. The FBI estimated the revenue from the trades on Silk Road over the 2.5 years of the site's operation in the order of $1.2 billion. Bitcoin became the currency of choice for the parties to these illicit transactions, attracting them with its perceived anonymity and operations outside of the legal system. On October 2, 2013, US law enforcement shut down Silk Road and arrested Ross William Ulbricht, who in 2015 was convicted of running the site and sentenced to life in prison. In the process, the FBI seized about 26,000 bitcoins, then worth approximately $3.5 million.

All these events, inevitably, attracted regulators' and policymakers' attention to Bitcoin. In the US, Senate hearings were held on Bitcoin on

November 18–19, 2013. The digital currency made a generally positive impression, and even though policymakers stressed its potential risks, no immediate regulation was recommended. In some other countries, the reaction was harsher. China's central bank, likely still remembering the Q-coin episode we described in the previous chapter, banned financial institutions from handling the cryptocurrency. Consequently, the Baidu website stopped accepting bitcoins in December. Similarly, Vietnam's financial authorities made the currency outright illegal in that country.

In the presence of these and similar stories, Bitcoin made it to the mainstream news. Even though it lacked details, the general public heard about this "Bitcoin"—an emerging digital currency, with no central bank, defying national borders, which was gaining in value and popularity. Back in 2015, Bitcoin was touted as an instantaneous, anonymous and free way to make transactions. It started to be perceived as a quicker and cheaper alternative to existing money, to be used in peer-to-peer transactions, international transfers, etc. As we will see, at least some of that enthusiasm was misplaced. It turns out that paying with bitcoins is not completely anonymous and it is rarely free or instantaneous. The revolution that Bitcoin was expected to bring did not occur—the new form of money did not replace the old one.

Bitcoin nevertheless proved resilient. It has a consistent recurring presence in the media, with very similar themes as in 2013: illegal activity, price volatility, and interest of regulators. After the FBI shut down Silk Road, many more darknet marketplaces for illegal trade have appeared, with Bitcoin as the default payment mechanism. Moreover, ransomware and ransom attacks have become ubiquitous, with businesses and households alike being forced to pay the ransom in bitcoins to regain access to their files.

While the $1,200 price at the end of 2013 seemed mindboggling, it paled with almost $20,000 it reached at the end of 2017, which was dwarfed again by over $60,000 in 2021. Between these surges, Bitcoin's price dropped rapidly. For example, Bitcoin nearly doubled and then lost half of its value in just four months early 2021. The significant price volatility is also present in shorter terms. It is not unusual for the cryptocurrency to experience a price change of 10% within the same day. It is very likely that this volatility is due to Bitcoin being again a buzzword, helped by stances and public messages by celebrities like Elon Musk, but also because of renewed attention by regulators. The year 2021 has been

marked, for instance, by El Salvador giving Bitcoin the status of legal tender and China's crackdown on "Bitcoin mining."

Bitcoin still inspires enthusiasm, albeit different than in 2013. It is no longer expected to be making transactions cheap, fast, or anonymous. Now, it is expected to be a new type of investment asset. The hopes put in Bitcoin changed even though Bitcoin's protocol and technological capabilities remain constant. This is because the novelty of a cryptocurrency can capture attention and create enthusiasm even though many people do not understand how it works. This allows for some misleading statements and promises to persist in the common discourse.

Despite all the misunderstandings, Bitcoin is an ingenious development in computer science. Its major contribution, which goes beyond its potential use as a currency, is that it solves the double-spending problem in a permissionless decentralized network.

4.1 The Double-Spending Problem

The double-spending problem was the major stumbling block, for a long time perceived to be an insurmountable obstacle in the development of decentralized digital currencies. To illustrate its nature, we will begin with a simple thought experiment.

Suppose you had a technology that would allow you to perfectly copy money, say an ingenious photocopying machine that could quickly and easily duplicate banknotes. In Chapter 2 we mentioned counterfeiting traditional money; here we are talking about creating copies that would be absolutely indistinguishable from the originals.

If you were the only person with access to such technology, you might enjoy it for a while (we note that using it would, of course, be illegal—which is why we're keeping this discussion limited to a *thought* experiment). If instead this copying technology were widespread, nobody would care to work to earn money. Why bother with a job if you could simply copy the money you need? As long as you have a unit of money to start with, you can double spend it, or even triple- and multiple-spend it. It suffices to copy it and multiply the original as much as you wish. However, at the same time, nobody would want to sell anything to another person—why part with an object or a service if what I'm getting in return is something I could have replicated myself in the first place?

In other words, money would cease to function and the economy would grind to a halt, unless it switched to a different, more difficult

to copy, currency. This simple example illustrates that something that is easy to copy would not make very good money.

All this brings us to digital currency. Digital currency is essentially a string of zeros and ones, perhaps encoded on a magnetic strip, on a chip, or stored somewhere in the cloud. Regardless of where it sits, this piece of data is imminently copy-able: we can reproduce it exactly, in as many copies as we wish, without harming the original. If money were simply electronic impulses, it seems we would be perilously close to the thought experiment above.

Perhaps the easiest solution is to keep a ledger, an account that would list each unit of the digital currency (perhaps by its serial number) and keep track of who owns that unit at any given time. After a transaction, the ledger would be updated by changing the ownership of the currency unit from the buyer to the seller.

Keeping such a ledger is a good idea, but we have not yet solved the problem completely. After all, a ledger in the digital world is just a piece of data, and one can copy it as easily as before. For example, a dishonest buyer may copy the ledger prior to a transaction. While the ledger would be updated in any transaction, the dishonest buyer would try to revert to its prior version that still lists him as the owner of a unit of currency he has just spent. That dishonest buyer could then be able to spend twice his/her coins, that is, to *double spend*. So, it seems we have merely replaced the problem of copying the digital currency with the problem of maintaining the integrity of the ledger.

Things would be different if we could designate a trusted third party that would be in charge of the ledger. The digital currency would then be centralized in a sense that the trusted party would be the only party with the right to alter the ledger, and would diligently and truthfully record all transactions in it. All transactions would need to be reported to that trusted party, and sellers would consult it to verify that a prospective buyer has enough funds to complete a transaction.

Digital currencies managed in such a centralized fashion would and in fact do work. This is what banks do when they keep our deposit accounts or credit card accounts. All platform-based currencies we discussed in the previous chapter are also organized this way. Whether we talk about Amazon Coins or Facebook Credits, there is always an institution in the background that keeps track of all accounts and that stands ready to update the records whenever a transaction occurs. This institution has

information about everybody's holdings and about all transactions that take place. This is very different from the anonymity of cash transactions.

Is it possible to design a decentralized digital currency that could operate as money with no intermediary, that is, even though there would be no centralized party to keep track of the transactions? Initially, the consensus among computer scientists was that this would be difficult or perhaps just impossible. In fact, the e-cash problem has been a long-standing challenge in computer science since the early 1980s. The solution to this puzzle was finally proposed in 2008 in a paper published by Satoshi Nakamoto, "Bitcoin: A Peer-to-peer Electronic Cash System."

The impact of Nakamoto's paper has been immense. The solution he (or she, or they—we still do not know who is behind the pseudonym) proposed, known as the Bitcoin protocol, was the first well-working solution to the problem of decentralized digital currency. More precisely, Bitcoin was the first fully functional decentralized solution to the problem of double-spending discussed above. As such, it is an important contribution to cryptography and to computer science in general. Moreover, as we will see later in this chapter, multiple hundreds of such cryptocurrencies have been proposed. While they differ on a number of dimensions, many of them share the reliance on the same general technology as Bitcoin does. All these currencies, including Bitcoin, are commonly referred to as cryptocurrencies, to reflect the idea that the soundness of the system depends only on the algorithm and cryptographic tools.

4.2 A Brief Overview of Bitcoin's Design

We will limit our discussion of how Bitcoin works to a high-level overview that avoids some of the technical intricacies and the computer-scientific innovation Bitcoin is justly famous for.[1] Our intention is not to give a detailed description of the inner workings of Bitcoin, but rather to illustrate the mechanism, and especially the incentives involved. We will try to be technical only inasmuch as it contributes to a better appreciation of the economic forces affecting the system.

The most important innovation of Bitcoin was to create a digital currency system where anyone can participate in maintaining the ledger of transactions, and this ledger is resistant to malicious changes, like double

[1] There are many excellent sources explaining technical aspects of Bitcoin, for example Antonopoulos (2014) or Haeringer and Halaburda (2018).

spending. With that, the system is the first one to allow for a reliable digital currency without the need of a trusted party to safeguard it—no trusted party maintains the ledger and no party needs to grant permissions to participate in the system. In this sense, Bitcoin is the first *permissionless* decentralized digital currency system. Such a system is achieved by a combination of cryptographic tools and economic incentives.

Following a common convention, we refer to the concept and the system as Bitcoin, with capital "B," reserving the lower-case "bitcoin" for the currency units. All transactions involving bitcoins are written on a transparent ledger, so that at any given time anyone can trace the history and confirm that a given bitcoin or part thereof is not being double spent. Despite transparency, the ledger preserves privacy to some degree, as the parties to the transactions are not identified by name but by the addresses used in the transactions. An address is a number represented by a string of 26 to 35 alphanumeric characters. Since the numbers are large, they are often represented in hexadecimal system, which uses letters to help represent numbers in a smaller space.[2] There is no limit on how many new addresses a user can obtain. Thus, it is possible to use a different address for every receiving transaction.

The addresses rely on a standard computer scientific concept of public key encryption. Public key encryption is commonly applied in many internet systems, like email or login passwords. Keys are created in pairs: a matching public and private key. A Bitcoin address is derived from the public key, and is designed to be freely shared. When a person wants to pay with bitcoins, he or she sends (i.e., broadcasts to the network) the transaction, which includes the sender's signature. The signature is based on the sender's private key along with the information included in the transaction. Because it is mathematically related to the sender's address, it proves that the sender has the right to spend the bitcoins that have been received by this address. The way the public key protocol is designed permits anyone to check that the transaction was signed using the private key associated with the public key (or the address) of the sender, without needing to know the private key. Since the sender's private key is the only

[2] The hexadecimal system uses base of 16 for number notation, similarly to the decimal system using the base of 10. That means that each digit may represent any number between 0 and 15. Since we do not have commonly accepted digits for 10, 11, …, 15, the hexadecimal system may use, for example, a, b, …, f for that.

thing that is needed to create a valid signature, aside from public transaction information, owners of bitcoins are well advised to keep their private key secret. Otherwise, anyone who knows the private key can use the bitcoins in the address this key controls. Also, if someone else spends your bitcoins (or if you lose your private key) there is little hope of regaining control of the address, as there is no customer support to call.

The importance of signing transactions is actually similar to what happens in a centralized network, say when one pays with Q-coins or Amazon coins, or from a bank account. There too you need to prove that you have the right to spend a given coin, although you do this in a different manner. When dealing with platform-based currencies or a bank, you identify yourself by logging into the platform, which keeps track of all holdings of digital currency in your account. When you transact with somebody, the platform checks that the funds are indeed available on the account, adjusts the balance of your account and the account you are transacting with. It also issues a confirmation that the funds were transferred.

The key innovation of Bitcoin is that such a trusted third party is no longer necessary. The signed transaction is broadcasted to the Bitcoin network consisting of miners. Anyone can become a miner by downloading a piece of software (that's at the core of the system being "permissionless"). Each miner—actually, the miner's software—collects the new transactions it has heard of into a block that needs to be added into the ledger. The transactions are verified by checking against the existing ledger that the bitcoins are sent by someone who has received them earlier (and properly signed), and that they have not been spent before. This verification is computationally easy. But we could not stop here.

If each miner creates a block independently, different miners could include different transactions, which would lead to different ledgers. For all the talk about "the ledger," as there is no trusted third party to keep one authoritative ledger, we instead have many miners each holding their own local ledgers which are updated based on the messages the miners send and receive to each other. Not every miner receives the same messages at the same time. The beauty of the system is that through its protocol the miners end up with exactly the same ledgers, i.e., they achieve a consensus. To coordinate and reach consensus on adding the same block of transactions to the existing ledger, the network needs to pick a block created by one miner that will be consistently added by

all miners. In order to choose whose block will become part of the commonly accepted ledger, miners draw a lottery of sorts, called the *proof-of-work*.

The lottery is based on a hashing function, which is a one-way function—given the input, it is easy to calculate the output, but there is no way to reverse engineer the input by knowing the output, or even part of the input. The output of a hashing function is a number represented as a string of characters of a fixed length, independently of the size of the input. Letters are used in the representation of the number if its written, e.g., in hexadecimal notation. A key characteristic of a cryptographic hashing function is that the output is extremely sensitive to any input alteration. For instance, the hash of the word "hello" is.

2cf24dba5fb0a30e26e83b2ac5b9e29e1b161e5c1fa7425e73043362938b9824

but if we instead write "hellO" (with an uppercase O) we get a very different hash,

04a6f55face2f46be8c23f627d539827615851e10751b63ec59db6d2c706b770.[3]

Aside from the transactions, every block includes the hash of the previous block, and a *nonce*—a number set by a miner. In the proof-of-work, the miners need to find a nonce such that the hash of the proposed block is below a certain number called *target*. The target number is in the same format as the hash output, but a smaller number, so it starts with some zeros. Since hashing is a one-way function, the only way to find a nonce that yields a hash below the target is trial-and-error. The lower the target (i.e., starts with more zeros), the more difficult it is to find a valid nonce. Hence lowering the target means increasing difficulty of mining. The algorithm in Bitcoin software automatically adjusts the difficulty every two weeks so that a valid nonce is found in the system every 10 minutes on average. Such adjustment is needed to account for the varying number of miners and the total amount of computing power, as well as technological progress in the speed of computing.

Once a miner has found the nonce for his block, he broadcasts the block to the network. All the other miners, upon receiving the broadcast

[3] There exist different hashing algorithms. The one we used here is called SHA-256, which is the same as Bitcoin's.

and checking validity of the broadcasted block, add it to their ledgers. The hash of this new block then becomes "the hash of the previous block," and since miners now have to use this new "previous hash," all the work they have done before is now irrelevant. This is so even the block a miner was working on does not contain any of the transactions that are included in the newly broadcasted block.

While calculating one hash is computationally trivial, finding a valid nonce may require calculating thousands of hashes. That quickly becomes computational-resource intensive. The miners are incentivized to participate in this costly process by mining rewards collected by successful miners. The mining reward consists of new bitcoins, called *block reward* or *coinbase*, and transaction fees. This is the only way new bitcoins are created. Thus, the mining activity does not only consist of processing and recording transactions but also of supplying new bitcoins to the system.

The schedule of new bitcoins created with each block is fixed by the Bitcoin protocol. It was initially set at 50 bitcoins per block, and is halved every 210,000 blocks (approximately every 4 years). The block reward was halved to 25 bitcoins on November 28, 2013, then to 12.5 bitcoins on July 9, 2016, and recently to 6.25 bitcoins on May 11, 2020. The next halving is expected to occur in 2024. Eventually this halving process will reduce the block reward to one satoshi (the smallest denomination in the Bitcoin system, equal to one hundred millionth of a bitcoin, i.e., 0.00000001 bitcoin) and, approximately in 2140, should Bitcoin still exist, the coinbase will disappear. By then, the total amount of all minted bitcoins will be a tad below 21 million. This design decision was motivated by the desire to assure scarcity of bitcoins (roughly speaking, make it similar to gold), but, as we will see later, this may create undesired deflationary consequences for the Bitcoin economy.

Once Bitcoin reaches its fixed supply, there will be no new bitcoins to provide the incentive to participate. Instead, miners will be compensated only with transaction fees. Fees are voluntarily added by the sender of a transaction. The fee is collected by the miner who successfully adds this particular transaction in his block to the ledger, along with the block reward.

While the senders of the transactions do not need to pay any fees, they have incentives to do so. Since miners are free to select the transactions they want to include in their block, transactions that carry the highest fee will be selected first. Therefore, adding fees increases the probability that the transaction will be verified and added to the blockchain sooner. The

level of fees varies over time. Until 2017, the fees were on the order of a few cents, and many transactions were processed without any fee. In contrast, for the first half of 2021 the average fee was oscillating around $15, rising to $60 in April 2021.

The transaction fees are driven by the demand for transactions to be processed and the fixed capacity of Bitcoin blocks.[4] A block cannot exceed 1 MB, which could hold at most 4000 transactions. Since a block is processed on average every 10 minutes, a demand to process more transactions than the capacity leads users to compete with each other for the limited block space by bidding up the transaction fees.

Since only one block at a time is added to the ledger, only one miner collects the block reward and the transaction fees from that block—the miner "winning" the hash lottery of mining—even though all miners use their computational resources. This creates a tournament structure for the miners. They compete one against another and the reward they earn is all-or-nothing. The probability that a miner wins the contest is proportional to its share of the computing power in the Bitcoin network. Such structure incentivizes the miners to invest in more computational power, to be able to compute more hashes than their competitors within the same time, akin to buying more lottery tickets.

How much they are willing to invest depends on the value of the reward, and so on the price of Bitcoin. If the price of Bitcoin increases, the value of block reward increases as well, and miners find it worthwhile to invest in additional computational power to increase chances of winning the reward. Ironically, if everyone invests proportionally, the odds of winning stay the same. But if everyone else invests, a miner needs to invest too lest his odds of winning fall. Thus, a higher price of Bitcoin will incentivize existing users to increase the mining power and will also attract new miners—increasing the cost of mining. In fact, because of free entry into mining, the marginal miner will break even, i.e., will find his cost of mining equal to the expected reward net of his opportunity.[5] Miners are not equal—some are likely to face lower cost of electricity, cooling,

[4] See Huberman et al. (2021).

[5] This condition is called by economists a "zero profit condition" and is a standard tool in economics to analyze market entry and exit. It refers to the *economic* profit, which is equal to the *accounting* profit (revenue minus cost) minus the opportunity cost. A miner's opportunity cost for mining is simply the revenue the miner would get if he were not mining. Hence, miners will enter the mining game as long as their accounting profit

or equipment and thus earn positive profits. Miners thus have an incentive to find ways to reduce the cost (electricity or equipment) or make mining more efficient (yield more computing power for the same electricity consumption). Miners who manage to do so will see an increase in their economic profit. Competitive forces that are at the heart of the mining activity thus create a never-ending arms race into mining power.

Initially, Bitcoin was mined on CPUs, i.e., regular computers.[6] But nowadays, the investment in finding a valid nonce is not inconsequential: while in principle anyone can start mining even on a regular computer, becoming a meaningful miner in the Bitcoin network requires a fixed investment in the hardware (e.g., application specific integrated circuit—ASIC—machines designed to focus on particular operations, in this case, the Bitcoin hashing function) and considerable amounts of electricity. That last element is important enough for serious miners to locate in places where the cost of electricity and of cooling their machines is low, for example, in Iceland, or near an abundant source of water, at the risk of creating serious environmental concerns. For instance, the installation of mining rigs near the Seneca Lake—part of the Finger Lakes in New York State—is a cause of concern as the water, used to cool the rigs, is now warming up the lake, affecting local fauna and plants.[7]

Proof-of-work is a costly way to run the lottery among miners to decide whose block will be added to the ledger. The main role of this cost is to facilitate immutability of the ledger. Every block of transactions includes the hash of the previous block. This reference to the previous block links the blocks into a chain and hence the ledger created by the Bitcoin protocol is often called *blockchain*.[8] Including the hash of the previous block allows anyone to check if there is any alteration in the blockchain. Such a set up involving hash links was originally proposed by

is at least as high as their opportunity cost, i.e., as long as their economic is positive. See Prat and Walter (2021).

[6] One of Satoshi Nakamoto's intentions stated in his 2008 whitepaper was to assure one-CPU-one-vote, as opposed to one-IP address-one-vote. Nakamoto worried that one CPU can take over several IP addresses and have more than democratic power.

[7] https://www.nbcnews.com/science/environment/some-locals-say-bitcoin-mining-operation-ruining-one-finger-lakes-n1272938.

[8] This linkage of blocks of data is a defining feature of any blockchain, independently of its accompanying protocol or its purpose. It therefore may be misleading to expect that any blockchain will have the same properties as Bitcoin's blockchain.

Haber and Stornetta in the early 1990s in the context of time-stamping digital documents.[9]

Suppose you wanted to go back and change a transaction, for example, replacing the recipient of the bitcoins with yourself. Modifying or removing a transaction in one of the past blocks will cause the hash of that block to change and no longer coincide with the hash contained in the next block.[10] Carrying the change further changes the hashes of all subsequent blocks and most likely makes them above the target. Since the target starts with many zeros—and hence valid block hashes need to start with at least the same number of zeros—it would be clearly visible to anyone that something was retrospectively changed in the ledger. This makes the Bitcoin blockchain *tamper evident*.

To cover the evidence of tampering, an attacker would need to find a new nonce that would yield exactly the same block hash. By the property of the hashing function used in Bitcoin, it is easier to find one particular grain of sand in the whole world than to find such a nonce. An alternative way to cover the evidence of tampering is to redo all the hashes so that they are different but below the target. Redoing the hashes is as costly as mining. Thus, the more costly mining is, the more costly it is for the attacker to cover the tracks of tampering. And since the cost of mining depends on the value of the mining reward—the higher the price of Bitcoin, the safer the blockchain is.

Now, let's suppose an attacker has borne the cost and has succeeded in creating an alternative internally consistent blockchain in the sense that there is no double spending recorded in the ledger and the hashes check out; and now broadcasts this ledger to other miners. Recall that every miner keeps its own local blockchain, with the consensus mechanism ensuring that all those local ledgers are the same. Other miners now see two different blockchains. We call such a situation a *fork* in the blockchain. It would be easy to say that they reject the blockchain that they have received later—that would prevent any alterations even if the attacker succeeds in redoing the proof-of-work.

Such a strategy of sticking to older blockchain could bring serious problems, as it could create a permanent split in the network where two or

[9] Haber and Stornetta (1990).

[10] Each block also contains a series of hashes (called a *Merkle Tree*) that serves a similar purpose but at the transaction level.

more parts of the network do not agree on the version of the ledger—the local versions of the ledger held by the miners differ, and they know they differ. In other words, the consensus would fall apart, and such different local ledgers could not function as a currency ledger. The reason why sticking to the older blockchain would be a problem is that forks can happen accidentally on the Bitcoin network even if all the miners follow the protocol honestly. Due to the peer-to-peer network structure and stochastic nature of the mining process it is possible that two or more miners find and broadcast a new block at approximately the same time (that happens frequently). In such a case, each miner would consider the first block it received as the legitimate one and reject the others blocks. Since not all miners would receive in the same order those "simultaneous" blocks, miners will have different versions of the blockchain. A protocol is thus required to restore consensus by selecting a unique blockchain whenever multiple ones were created in a fork.

Consensus is achieved with the following rule: when miners are facing competing internally consistent versions of the blockchain they are advised to focus on the longest one, that is, the version that has the larger number of blocks. Since the pace at which blocks are added is not constant (10 minutes is only an average), sooner or later one of the versions will be longer and thus become a focal point for miners. The other versions will become defunct. Note that there is no obligation for miners to follow this *longest chain rule*, but under the expectation that all the other miners do so, it is in the best interest of each miner to do it as well.

Combining proof-of-work and the longest chain rule is Nakamoto's key contribution. It is the combination of these two features that yields consensus and facilitates consistency of the blockchain over time. The attacker needs to not only find new nonces to the existing blocks, but needs to create a longer blockchain for the attack to succeed. The mining difficulty is the same for the attacker as for the other miners. This implies that an attacker, being alone working on his version of the blockchain, will manage to add new blocks at a slower pace than the rest of the miners. In the Bitcoin whitepaper, Nakamoto provides calculations showing that the probability of a successful attack rapidly converges to zero as the computing power in the network increases. An attack is thus guaranteed to succeed only if the attacker has more than 50% of the computing power of the whole network. Gaining such computational power is very costly, which was the intent in Bitcoin network design.

Standard economic analysis shows that incentives in organizations usually take the form of rewards and penalties, which in general crucially rely on knowing participants' identities and having a central, policing authority. Without these conditions, one would assume that a certain level of trust and the presence of well-intentioned participants are needed to achieve the collective good. In that respect, Bitcoin's design, which assumes de facto that participants (miners) do not know and do not trust each other, is an extraordinary feat. The key in Bitcoin's design is to make any deviation extremely costly by making any fork of the blockchain unlikely to be successful. This is achieved by linking all miners together through the difficulty parameter. That parameter does not affect the pace at which the blockchain grows (on average one block every 10 minutes) but that of any individual fork, which is condemned to grow at a slower pace as soon as miners follow the longest chain rule.

For economists, Nakamoto's design is thus reminiscent of Adam Smith's *invisible hand* interpretation of the market reaching an equilibrium through economic forces. Economic agents or miners, solely driven by their own personal benefit, can lead the system to reach an optimal state: a supply–demand equilibrium in case of Adam Smith's analysis of markets, and a censorship resistant consensus in Nakamoto's cryptocurrency.

4.3 Not the First One—Predecessors of Bitcoin

Our description of how Bitcoin works focused on the essential parts, so we can focus on economic forces and competition later in this chapter. But even from this simplified description one can see that it is very demanding to construct a decentralized currency system that solves the double-spending problem. In fact, it took many attempts to do so. Bitcoin was not the first decentralized digital currency—although it was the first one that worked well enough to gain some acceptance by the general public. In its design, Bitcoin incorporated many of the earlier solutions. The cryptography community has been interested in developing a decentralized currency system since the rise of the Internet.

The first piece of Bitcoin-like technology was *hashcash*, a system based on proof-of-work introduced in 1997 by Adam Beck. Beck's purpose was to prevent email spam by requiring the sender's computer to do computational work before sending the email. Such work would be relatively trivial for an individual email and would not affect computer performance

for that use. However, it would make sending thousands or millions of emails prohibitively costly in terms of computing power, making sending mass spam emails uneconomical. The ingenuity of hashcash is that it obtained this goal without charging money for emails. As we saw, Satoshi Nakamoto incorporated this element into Bitcoin to make it costly to create a fake blockchain.

In 1998 Wei Dai designed a decentralized digital currency called *b-money* that would allow for anonymous peer-to-peer transactions. The transactions would be recorded by the members of the network in a ledger. Each participant would have a copy of the ledger. To fight misconduct (e.g., recording transactions that did not happen), the nodes in the system had to deposit money to a common pool, which was used for fines for misconduct and rewards for proof of misconduct. Such a system of fines and rewards, however, is difficult (but maybe not impossible) to enforce without a central authority to decide and solve disagreements.

In 2005 Nick Szabo proposed *bit-gold*, which also used proof-of-work and a distributed *property title registry* (similar to Bitcoin's ledger).[11] The work of solving a problem by trial-and-error (also similar to Bitcoin's mining) was used to create new pieces of bit-gold, but there was no clear control over how much bit-gold can be created, and how quickly. Szabo himself raised a concern that a powerful computer could "swamp the market with bit gold," lowering its value because the market will adjust.[12]

B-money and bit-gold were ideas, theoretical considerations, never really implemented, making it difficult to know how well they would work. They had never captured enough interest from people outside the small group of cryptography enthusiasts. But there were also commercial efforts to create anonymous digital currency systems. Similar to Bitcoin, these systems comprised independent currency units, allowed for greater divisibility, and involved a universal permanent ledger of transactions. However, those systems were centralized. The two most prominent examples are DigiCash and Citibank's e-cash called Electronic Monetary System.

[11] See Nick Szabo's blog entry about bit-gold: http://unenumerated.blogspot.ca/2005/12/bit-gold.html.

[12] In the context of traditional currencies, an analogy here would be Spain mining gold and silver in the Americas, dramatically increasing the supply of those metals in Europe and lowering their value.

DigiCash was a commercial company, set up in 1989 by David Chaum, and it proposed building a system of anonymous electronic cash to governments and banks. The DigiCash system had asymmetric anonymity: the payer was anonymous, while the payee could be "irrefutably identified if needed." This feature was motivated by the desire to end corruption and organized crime. The innovation of the system was the ability to transport information wirelessly, and thus it was well suited to pay road tolls, which was supposed to be its first use. David Chaum had even signed a contract with the Dutch government for this purpose. The idea of the DigiCash system also attracted some interest beyond toll application. There was interest from banks (like Deutsche Bank and Credit Suisse), Visa and Microsoft. But nothing came out of it, and by the end of the 1990s everything fell apart, including the company itself. For a few years, one bank in the US, The Mark Twain Bank of St. Louis, MO was using DigiCash. But that ended in 1997.

The second example of commercial development of a decentralized digital currency was *Citibank's e-cash*. In the 1990's Citibank was developing an in-house system of electronic money. The money had the interesting feature that it expired after some time, and the holder needed to contact the bank to replace it. This feature was meant to prevent money laundering. There were test runs and pilot programs in 1997 and 2001. In 2001 the project was shut down by the new management of Citigroup.[13]

Bitcoin took some elements of these earlier systems and combined them in a new form. That new system includes elements that had been by then common and expected, for example, its peer-to-peer nature (anyone with a computer could become part of the network) or its use of the public-key encryption with private key. Its novelty and importance came from combining the idea of a blockchain—a public ledger that would be prohibitively costly to forge due to proof-of-work—and mining—the monetary incentive system to encourage the nodes to keep the ledger up to date. Those two features make it possible to keep the system honest while fighting off hackers.

[13] Vigna, Paul and Michael J. Casey (2015). See also http://archive.wired.com/wired/archive/2.12/emoney_pr.html.

4.4 New Challenges Created by Bitcoin Design

Bitcoin is the first system that achieved the goal of sustaining reliable decentralized digital currency. In spite of all its ingenuity, Bitcoin is not without shortcomings. The substantial and still increasing cost of mining is certainly the most evident issue, but there are other problems that are no less important, such as Bitcoin's scalability or throughput problem.

Implicit in Bitcoin's design is the time dimension: a blockchain growing too fast would necessarily increase the number of occurrences of accidental forks—and thus opportunities to conduct an attack. Similarly, the larger the blocks the more difficult for miners to efficiently communicate with each other the creation of new blocks. For this reason, Nakamoto set the block's maximum size to 1 MB, corresponding to about 4000 simple Bitcoin transactions (i.e., transactions with only one input and one output). With an average delay of 10 minutes between blocks this entails a maximum throughput of approximately only 7 transactions per second! This clearly pales in comparison to Visa or Mastercard, which are believed to be able to handle several tens of thousands of transactions per second.[14]

There are also other problems, perhaps less evident at first sight but not less important, such as pool mining (which can fragilize the system), the existence of types of attacks that were initially ignored or Bitcoin's governance structure (or lack thereof) that makes improvements hard to implement.

4.4.1 Mining Arms Race and Electricity Consumption

The most obvious part of the mining cost is the electricity. Moreover, one needs significant investment to be competitive in the mining business. It is no longer enough to mine on a computer or even a cluster of computers, one needs specialized mining rigs designed to run the Bitcoin's hashing function, SHA-256, as efficiently as possible in order to find nonces more quickly.

We see an arms race in the mining business, with miners continually investing in new hardware to build a competitive edge, and pushing

[14] Visa handled about 188 billion transactions in 2020, which gives an average of nearly 6,000 transactions per second (https://www.statista.com/statistics/261327/number-of-per-card-credit-card-transactions-worldwide-by-brand-as-of-2011/). In its press releases Visa, Inc claims to have a capacity of 65,000 transactions per second (e.g., https://usa.visa.com/about-visa/newsroom/press-releases.releaseId.17706.html).

their competitors to do the same. Initially, bitcoins were mined with regular computers. Eventually, one of the early miners noticed that the graphics card attains computational advantage in mining. This gave rise to designing devices that would be ever more efficient in running SHA-256, but not useful for any other calculations.

This ruthless race towards new and more powerful technology arises because of the tournament structure of the Bitcoin algorithm. Since the winner of the mining lottery takes the whole reward, even slight improvements that put a miner just a bit ahead of everyone else give the miner a large expected reward. At any given point, the incremental investment may seem small and be worthwhile, but when everyone else also invests in response and catches up, the total investment of the overall mining industry may easily become worth more than the value that miners can win.

The race is sped up by a particular feature of the Bitcoin system: the difficulty of finding a valid nonce is adjusted to keep the expansion of the blockchain to a constant pace of one block being added every 10 minutes. With more computing power finding an acceptable nonce takes less time. To slow miners down so as to keep the 10-minutes lag constant, the algorithm increases the mining difficulty, requiring more hashing operations to be conducted on average before finding a valid nonce. A side effect of the difficulty increase is a higher energy use: even though new mining rigs are designed to operate more efficiently, running more computations typically requires more electricity.

The rising energy costs impose an externality on the environment and on the overall economy. By design, only one computation of the miner who wins the race is important in a sense that it leads to adding a new block to the blockchain. All other computations for the winning and for the other miners are discarded: the addition of a new block changed the "hash of the last block," meaning that the computations that have been done so far are of no use for the next block. It is in that sense that the energy spent on the discarded computations is a loss to the system. It is worth keeping in mind, however, that the energy spent by the miners who lost the competition is not exactly lost: that "lost" energy spending is directly linked to the mining difficulty, which makes hacking or tampering the blockchain difficult. In other words, calculations that may be considered as futile are the very determinants of Bitcoin's security.

Bitcoin's energy consumption and the resulting environmental impact is one of its main drawbacks and is a reason for concern. According to

Cambridge University's Centre for Alternate Finance (CCAF), the Bitcoin network was consuming nearly 150 terawatt hours in April 2010.[15] This is the equivalent to the electricity consumption of Poland (the 24th country in terms of electricity consumption). Such assessments are obtained by estimating the computing power needed to find a valid nonce and the electricity consumption of the average mining rig. But there is considerable uncertainty regarding the exact number of mining rigs and their respective computing power. The CCAF acknowledges this problem by explaining that the consumption power could be in fact anywhere between 46 and 500 terawatt hours (Portugal, 52nd, and South Korea, 9th).

Since most of the world's electricity is produced burning fossil fuel, high levels of electricity consumption are often associated with high levels of carbon footprint. But estimates about Bitcoin's carbon footprint are even harder to pin down: electricity consumption is not equivalent to carbon emission. Using geographical data about miners the CCAF estimated in September 2020 that about 39% of the electricity used by miners come from renewable energy. At the same time, CoinShares, a European fund specialized in crypto-assets, estimated it to be 73%.[16] While those estimates certainly suggest that a significant proportion of Bitcoin mining is using renewable energies (and thus mitigating Bitcoin's carbon footprint), the sheer difference between those estimates also signals that the exercise is far from being trivial.

Moreover, ascertaining Bitcoin's energy impact is further complicated by the fact that Bitcoin can be mined anywhere—it does not need to be close to the end users. Miners can thus adapt by shutting down their mining farms where electricity is costly and turning on other mining farms where electricity is cheaper, if not free. For example, in the Sichuan and Yunnan provinces in China, the large excess of hydro-electric production during the wet season attracts a large proportion of the mining activity. The CCAF estimates that these two provinces account for 50% of the global mining activity during the wet season but only 10% during the dry season.[17]

[15] https://digiconomist.net/bitcoin-energy-consumption.

[16] Blandin et al. (2020).

[17] https://hbr.org/2021/05/how-much-energy-does-bitcoin-actually-consume.

While the carbon footprint of Bitcoin mining imposes a global externality, the related electricity consumption also imposes local externality. Bitcoin mining facilities located in small communities consume such a substantial fraction of energy that the local energy price affects everyone in the area. At the same time, the mining facilities do not create more jobs nor add to the local infrastructure. For instance, for Upstate New York in the USA it is estimated that Bitcoin mining increased the electricity bill by about $79 million for small businesses and $165 million for households, an amount that is only partially offset by an increase of tax revenue from mining.[18]

4.4.2 Mining Pools and Centralization Propensity

The mining arms race also increases the propensity to mining centralization. First, the arms race forces less efficient miners, or miners who cannot afford costly improvements to their mining rig, out of the system; even if those miners stay in the network, they will have a relatively lower share of the total computing power. That is, the high cost of mining equipment and frequent need for additional investments combined with uncertain payout increase barriers to entry. With fewer participants, it becomes more likely that one or few miners with considerable computational power will dominate the network.

Moreover, the arms race together with Bitcoin's winner-take-all tournament structure gives miners incentives to pool their resources into mining pools. Mining pools are co-ops of miners, who divide the mining tasks among themselves and share the rewards—typically proportionally to the computing power contributed to the pool. For individual miners, the incentive to get into the pool is to lower the uncertainty of the revenue stream. Winning the competition is profitable, but very unlikely for an individual miner in a limited period of time. Instead, participating in a pool allows users to share the risk and essentially insure one another. The pool will win more frequently than any individual alone, although of course the win brings a lower reward when the pool wins (as the newly minted bitcoins and fees earned need to be spread across the whole pool). Thus, for many miners, especially those with less mining power or that are risk averse, this tradeoff is attractive as it allows them to smooth their

[18] Benetton et al. (2021).

earnings over time. That is, they prefer to forego the possibility of large but infrequent prizes for the prospect of a steady accumulation of smaller rewards.

The intention behind Bitcoin was to develop a decentralized system. However, the existence of mining pools, where a large number of miners coordinate their mining efforts, leads to centralization. In mid-2021, just 6 pools—SlushPool, Poolin, F2Pool, ViaBTC, and AntPool—controlled more than 50% of the total computing power of the Bitcoin network. The relative importance of particular pools changes over time, but the dominance of a few pools has been an almost constant feature of the mining landscape for nearly all of Bitcoin's existence. Such structure affects the nature of the Bitcoin network because a pool is akin to a single but powerful miner. The presence of a few, yet large pools is thus tantamount to a mining game between just a few miners. In other words, the presence of mining pools increases concentration, which runs contrary to Bitcoin's decentralization objective. The irony is that such centralization forces are a natural consequence of the competitive environment set up by the Bitcoin protocol.

By aggregating the computing power of individual miners, it is even possible for a single pool to exceed 50% of the network's computing power. This is not just a theoretical possibility: several times in the past a pool has reached the 50% critical threshold. That happened, for instance, in June 2011 with the pool deepbit or in June 2014 with the pool Ghash.io.[19] One of the major innovations in Bitcoin was eliminating the need for a trusted third party who would monitor and manage the network. A miner or a mining pool that controls more than half of the network will essentially become such a third party dominating the network; ironically, it would not even be clear if such an entity may or may not be a "trusted" third party. Since such centralization propensity to a large extent comes from risk aversion, which is part of human nature—it is challenging to solve it with technology solutions.

4.4.3 Threat of Attacks

Centralization of mining is not just going against the ideals of Bitcoin. Such concentration of mining power may become an even more serious

[19] See messages posted in the bitcointalk.org forum, https://bit.ly/3kBdm6J and https://en.wikipedia.org/wiki/GHash.io.

challenge to Bitcoin because of its potential to lead to the "51% attack." The Bitcoin system maintains the integrity of the blockchain by relying on a diffuse network of miners who effectively keep each other honest. This system of distributed checks fails when a miner, or a coordinated group of miners, gains control over more than half of the computing power underlying the network. In such a case, the super-miner would be able to take control of the ledger, with powers ranging from preventing new transactions from being added to the blockchain, to potentially engaging in double spending.

Moreover, despite the name, an attacker does not need to control all 51% of the mining power in order to disrupt the blockchain. With just 33% of the mining power, an attacker already has a very high chance to succeed in rewriting the ledger by creating a longer chain, which constitutes a *longest chain attack*. With the majority of mining power, an attacker is guaranteed, if given enough time, to be able to create a longer chain.

The possibility of a 51% attack is not purely theoretical. In mid-2014, it was reported that Ghash.io, one of the largest Bitcoin mining pools, has briefly reached 50% of the computing power of the overall Bitcoin network.[20] No malicious behavior was observed. In following statements, Ghash.io was pointing out that mining pools have in fact no interest in attacking a system they profit from. Following that event, mining pools voluntarily adopted a policy of splitting up, should they reach a high proportion of mining power (around 20–30%). Inferring that this policy will increase confidence in the robustness of the system is a bit of a stretch, though, for splitting is not equivalent to decentralization; pools born from the split of a larger pool may still be run by the same or closely collaborating individuals. In fact, without any information about the exact relations between the individuals behind pools, it is not possible to gauge the extent of decentralization in the Bitcoin network.

Arguing that pools or miners would hesitate to attack as this could negatively affect the system's reputation (and thus the price of the coin) is actually contradicted by facts. Although there is no evidence that Bitcoin itself experienced a successful attack, it did happen for less pricey coins. The cryptocurrency *Bitcoin Gold*, a spin-off of Bitcoin, suffered two longest chain attacks, in May 2018 and in January 2020, with attackers

[20] http://www.coindesk.com/51-attacks-real-threat-bitcoin/.

being able to steal about $18 million and $72,000, respectively. Those attacks are not isolated cases. Another cryptocurrency, *Ethereum Classic*, also suffered a number of double-spending events via longest chain attacks in 2019 and 2020.[21] It is important to note that there may be many more attacks than the ones that were reported and spotted by specialists. Unless one keeps an archive of all the branches when a blockchain forks, the longest chain can only be spotted in real time when the blockchain is forked. Once the attack is successful, consensus on the blockchain is restored and thus there is only one version of the ledger. That is, if we do not observe a fork, we cannot say whether an attack occurred, and thus there is no evidence of data tampering. Those attacks on Bitcoin Gold and Ethereum Classic are even more interesting given that they have hardly affected their prices. This suggests an asymmetric relation between the price of a cryptocurrency and its security level: while a higher price does necessarily increase the security level of the system (attracting more miners and thus increasing the difficulty level), evidence of a security flaw might not affect the price.

These attacks show that one cannot take Bitcoin's security for granted, and more generally that, contrary to what many people believe, the protocol itself does not guarantee blockchain's security. Bitcoin Gold's and Ethereum Classic's protocols are very similar to Bitcoin's. Specifically, they use the same methods to prevent attacks. Thus, Bitcoin has the same vulnerability which the Bitcoin Gold and Ethereum Classic incidents demonstrated. As we have indicated previously in Sect. 4.2, Bitcoin's security is related to its price: the higher price, the higher the security level. This is so because a higher price also results in a higher value of mining reward, and thus a more intense competition between miners, resulting in more mining power, ultimately making mining more costly. The cost of an attack on Bitcoin thus increases with the mining cost. For the same reason, an attack on a less pricey cryptocurrency is cheaper, as such a cryptocurrency pays out smaller reward and therefore is computationally less demanding.

[21] See, for instance, https://arstechnica.com/information-technology/2019/01/almost-500000-in-ethereum-coin-stolen-by-forking-its-blockchain/ or https://www.bloomberg.com/news/articles/2019-01-08/ethereum-classic-movements-halted-by-coinbase-on-signs-of-attack.

4.4.4 Deflationary Pressure

Another weakness of Bitcoin as a currency is the potential deflationary pressure built into its algorithm. As we saw earlier, the Bitcoin supply (the number of bitcoins in existence) is increasing, but is doing so at a decreasing pace and at some stage will become fixed. This feature was consciously built into the design to prevent inflation, but may have unintended negative consequences. The scarcity may translate into downward pressure on prices denominated in Bitcoin (with fewer coins to go around, consumers may not want to spend too many coins on a given good).

Why would the limited supply of bitcoins translate into decreasing prices? To explain this phenomenon, we can use an economic theory called the "quantity theory of money." The theory links four economic quantities: the supply of money, M; the velocity of money, V (that is, how quickly money circulates in the economy); the goods and services the economy produces, Y; and the price of these products, P. These quantities are linked through an identity,

$$M\,V = P\,Y.$$

That identity is widely accepted among economists (after all, it is an identity), and has an appealing interpretation. The size of the economy (think GDP) is based on the number of goods and services that are being traded (Y) and on their prices (P). The sum total of these transactions needs to be supported by the money circulating in the economy. If money circulates very slowly (low velocity V) you need more of it to support the economy. For example, suppose that each unit of the currency, say, each separate dollar, can only be used once per year ($V = 1$). This means that to support the GDP of \$100 (the value of all goods and services equal to \$100), we need 100 separate dollars (or combination of separate banknotes and coins that add up to \$100).

The above identity helps us understand what happens when more goods are produced in the economy (that is, when Y increases). If the supply of money M is constant, and if the velocity of money V does not change, there is only one possibility: prices must drop. If they did not, we would not have enough money in the economy to support all the transactions that underlie the total production.

What does this theory predict for Bitcoin? First, note that as soon as the supply of Bitcoin is fixed, the supply of money M will be constant. If Bitcoin gains popularity and more people decide to use it, there will be

more products offered and purchased in the bitcoin economy—that is, Y will increase. The quantity theory of money tells us in response the level of prices, *P*, may drop proportionally. Simply put, there will not be enough bitcoins to support the increased spending, and, in response, prices will need to adjust.

Of course, a drop in prices is not inevitable. It may be that the fourth term of our equation, the velocity of money V, adjusts instead. If each bitcoin circulates in the economy faster than before, then the same supply of bitcoins will be able to support a larger volume of spending. Yet, unless Bitcoin experiences a major update, the Bitcoin velocity is unlikely to increase sufficiently to avoid a drop of the price levels: Bitcoin's current design allows for a maximum throughput of 7 transactions per second. A perhaps less attractive outcome would be a cap on the growth of the Bitcoin economy. If Bitcoin's use is limited to a relatively stable volume of goods and services (i.e., when Y above is fixed), then prices may not change even though money supply is constant. Either way, the identity above tells us that something has to give: it would be shortsighted to think that the size of the Bitcoin economy can change without having an impact on the level of prices.[22]

While falling prices may seem like a good thing, they tend to have an adverse effect on the economy. For example, people anticipating lower prices in the future will postpone their consumption and investments, which reduces the current size of the economy.

Given the above reasoning, why was it decided that the total supply of bitcoins will be constant? The likely reason was to build in an element of scarcity into the design of the cryptocurrency to ensure that it cannot be inflated. In the context of traditional currencies, inflation is often triggered by an increase in the supply of money.[23] The failsafe built into

[22] Our simplified analysis considers an economy that runs only on Bitcoin. The argument becomes more involved when the economy has two different currencies, say traditional money and a cryptocurrency. Still, even in that case the fixed supply of the cryptocurrency is likely to have a deflationary effect: as more people are trying to use the cryptocurrency to affect more transactions, the prices quoted in the cryptocurrency drop, and the exchange rate appreciates (the cryptocurrency becomes worth more units of the traditional currency).

[23] In terms of our identity, keeping the velocity of money constant, if there is more money to go around, but we have an unchanged number of goods, then prices of these goods need to adjust upward.

Bitcoin works so well; however, as to tilt the balance in the opposing direction and err on the side of deflation.

To offset the deflationary tendency, one may imagine introducing a gradual increase of the money supply into the Bitcoin algorithm. The problem then becomes getting the rate of increase exactly right, to ensure that prices remain relatively constant. It is doubtful (at best, debatable) whether there is a pre-specified formula that could achieve this goal; instead, in most countries, similar adjustments are left to central banks and are the outcomes of an ongoing thorough analysis depending on a number of economic variables and the state of the economy (crisis, growth, etc.). Judging from some of the narrative accompanying Bitcoin, at least some of its users are willing to accept the potential instability in prices in return for being independent of any institution such as a central bank. For these Bitcoin users, such a feature in Bitcoin's design would be perceived as positive and desirable.

4.4.5 Governance

Decentralization was a paramount objective when Bitcoin was designed. As we have seen in the context of mining pools, there are natural forces driving centralization instead. But even where it is achieved, decentralization also comes with drawbacks. One of such drawbacks of decentralization is the difficulty to adapt and enhance its design.

Over time, several limitations of the Bitcoin design became evident. In principle many of the issues such as limited throughput, deflationary trajectory, or maybe even excessive energy consumption could be fixed by updating the Bitcoin protocol. But decentralized governance makes any changes extremely challenging, if not impossible. There exists a well-organized community of developers working on maintaining Bitcoin and mining software updates are regularly made available, but such updates are in general about minor technical improvements such as bug correction, allowing users to use different encryption methods to sign their transactions, etc. More substantial modifications are more difficult to implement.

Updates to the Bitcoin software are made on a voluntary basis. Miners are free to update their systems following a proposal made by the developers, but are not obliged to do so. Updates that correct bugs in the software or make it more efficient are likely to be accepted by the vast majority of miners, if not all of them. It is more complicated for updates

that change the Bitcoin protocol or the structure of the blockchain, though. Such updates are difficult to implement because the blocks that will be mined with the new version of the Bitcoin software will be deemed as non-compliant under the old version. Those blocks will be rejected by the miners who haven't updated the software, and consequently the block rewards of those blocks will not be recognized by everyone. The number of miners (and their computing power) who announce their acceptance of an update is thus a key determinant of its success. The more miners do so, the more other miners have an incentive to also update their mining software.

An update being refused by a majority of miners may still be implemented, though. This is what happened to Bitcoin in 2017. At that time one of the most important debates in the Bitcoin community concerned the throughput of the system. A solution to this problem was proposed: the so-called *Segregated Witness* proposal (or *SegWit* as it is commonly referred to). This update, by changing the way the blocks are encoded, would allow for a larger number of transactions to be included in a block. A group of Bitcoin activists, developers and miners (mostly from China) opposed this update, favoring instead another proposal that would increase the blocks' maximum size. The disagreement between larger blocks and SegWit was more than purely technical—it could change the structure of the Bitcoin networks. The Bitcoin network not only consists of miners, it also contains nodes, which are relays to make the broadcast of new transactions and blocks faster and more reliable. Increasing the block size would make many nodes unable to operate efficiently and make the network more reliable on large third parties (that would host nodes) such as universities or private companies. As SegWit does not increase the size of the block, it would not affect the structure of the network.

Proponents of a higher block size agreed to update the mining software (allowing for larger blocks) should the SegWit update be activated, which eventually happened on July 21, 2017.[24] The update allowing for larger blocks was made effective shortly after, on August 1, which created a fork of the blockchain. One branch of the fork is compliant with the Segwit update, and the other branch compliant with the update allowing for larger blocks. The system that implemented the SegWit update kept the name Bitcoin, and the other system was named Bitcoin Cash. Both

[24] https://blog.bitmain.com/en/regarding-bitcoin-cash-viabtc-bitcoin-abc/.

Bitcoin and Bitcoin Cash have a blockchain that is common until the August 1, 2017 fork. So, transactions made before the fork are recognized by both Bitcoin and Bitcoin Cash, but not the transactions made after. About a year later, a disagreement within the Bitcoin Cash community led to another hard fork, with Bitcoin Cash splitting between Bitcoin Cash and Bitcoin SV (SV stands for Satoshi Vision). Bitcoin suffered another fork after the one with Bitcoin Cash in November, 2018, with the creation of Bitcoin Gold. This latter differs from Bitcoin in the hashing algorithm used for proof-of-work.

Forks like Bitcoin Cash, Bitcoin SV, or Bitcoin Gold are the consequence of the decentralized governance and may have an impact on the ecosystem of the cryptocurrencies. First, splitting the miners into multiple separate communities reduces the computing power of each network, thereby reducing the ability of the system to resist double spending via a longest chain attack. Such an outcome is not hypothetical, as discussed above: Bitcoin Gold suffered two successful attacks, in July 2018 and in January 2020. Second, forks create a de facto competition between cryptocurrencies. While users with holdings before the fork enjoy an increase of their holdings (e.g., holding 1 bitcoin before the August 1, 2017 fork meant holding 1 bitcoin and 1 bitcoin cash after the fork). Such hard forks can have negative or positive effects on prices. If the value of the coin depends substantially on the coin's security level, the price may drop because of a lower difficulty due to the mass of miners splitting between the two branches. But a hard fork also implies a higher diversity in terms of coin's design, which may in turn attract more users for either coin—a similar effect to the one observed with product differentiation.[25]

Bitcoin's decentralized governance is, to some extent, the biggest issue, as it makes correcting or updating Bitcoin's design to address its shortcomings a difficult, if not impossible task. Quite ironically, concentration of power—which runs contrary to Bitcoin's original motive—can compensate for the lack of centralized governance and help resolve governance standstill. A notable illustration of this is Bitcoin's fork in March 2013 following a technical update that did not roll out as expected. This fork created serious coordination problems among miners. Thanks to its large share of computing power, the pool BTC Guild managed to tilt

[25] Barrera and Hurder (2018).

momentum towards one of the branches, speeding up substantially the restoration of consensus.[26]

The inherent rigidity in Bitcoin's design actually motivated the creation of other cryptocurrencies aimed at alleviating (if not fixing) some of these problems. We will discuss this in the next chapter. In spite of its shortcomings, Bitcoin is a remarkable achievement: it is the very first fully functioning, decentralized digital currency. Nakamoto's goal was for Bitcoin to be used as money and rival cash. A natural question is how does Bitcoin compare with earlier forms of money in terms of usability.

4.5 How Do bitcoin's Attributes Compare to Earlier Money?

Since Bitcoin's design aimed to create a digital version of cash it is natural to ask how it compares to traditional currencies on their most important characteristics that we reviewed in Chapter 2. This is particularly relevant for any discussion of the competition between Bitcoin and such currencies, not just to answer the question of whether it is "better," but also to debate whether it is "good enough" to fulfill some or all of the functions traditional money serves today. While we consider these questions from the point of view of Bitcoin, the discussion here also applies to other cryptocurrencies, including those that attempted to fix some of Bitcoin's shortcomings.

We saw that one of the relevant characteristics of money is divisibility. Here, Bitcoin compares very favorably to traditional currencies, which typically operate using the metric system and are divisible up to a hundredth of a unit.[27] In contrast, Bitcoin allows precision to the eighth decimal place, with its smallest unit named "satoshi" after the inventor of

[26] A detailed report and analysis can be found here: https://freedom-to-tinker.com/2015/07/28/analyzing-the-2013-bitcoin-fork-centralized-decision-making-saved-the-day/.

[27] For accounting purposes, some prices may be posted in a fraction of a penny. But actual transactions are then always rounded up. In the US sometimes gas prices are posted at the gas stations with a tenth of a penny, e.g., $2.879 per gallon. This is because typically people buy many gallons. Even if the total comes up with a fraction, e.g., $31.669 for 11 gallons, it is rounded up to a cent, $31.67. Very small denominations of cash are often not available on principle. Canada is not issuing 1 cent coins anymore, because the value of the materials is larger than the nominal value of the coin. Cash transactions are rounded up to 5 cents. If the register rings CAD 13.22, you pay 13.20. If it rings 13.23,

the system. This provides for more divisibility and a higher precision not only than traditional currencies, but also than measuring barley or metal by weight. This enhanced divisibility may be useful for micropayments.

Another characteristic is durability, how long a currency can last. Again, the advantage here goes to Bitcoin. Bitcoins do not wear out or deteriorate. Of course, one can lose bitcoins. The media has reported a number of stories of people throwing away hard drives, or deleting wallets, and thus losing private keys that give them access to their bitcoins. Some estimate that about 20% of the bitcoins that have been mined so far—around 3.7million bitcoins—have been lost.[28] The bitcoins, however, are still on the blockchain, and will be there for as long as the Bitcoin network operates. From the point of view of the network, it is impossible to distinguish between a bitcoin that has been lost and a bitcoin the owner of which has not yet decided to spend it. When it comes to destroying money there is not much difference between Bitcoin and fiat money. You can lose a bill (or even a coin) permanently by destroying it or damaging it to the point that it is no longer recognizable. Similarly, bitcoins can be destroyed, or "burned" as it is commonly referred to. This can be done by sending bitcoins to an address for which no private key exists.

The bitcoins, being digital, are also easy to carry. There is of course the need for the software and hardware that manages them (e.g., a digital wallet on your smartphone). Is this easier or more difficult than carrying cash or a credit card? That may depend on the ease of use of the software and the hardware as well as on the individual preferences of users. A related question is the ease of transfer, which depends both on the available technology (access to computers, smartphones) and on the ecosystem (interface). When relying directly on the basic Bitcoin system, transfers are cumbersome. They are more difficult than handling cash (for person-to-person transactions) or than using credit cards for long-distance transactions. Indeed, private keys and addresses, which are necessary to submit a transaction, are made of long strings of characters that are difficult to memorize or prone to be mistyped. But such an argument can be easily dismissed, as it is only a matter of user interface. There are now

you pay 13.25. Interestingly, transactions with credit or debit cards are still with 1 cent precision.

[28] https://www.nytimes.com/2021/01/13/business/tens-of-billions-worth-of-bitcoin-have-been-locked-by-people-who-forgot-their-key.html.

many apps that make using Bitcoin as easy as any other payment service like Paypal or Venmo.

Storing bitcoins does not need to involve physical safes and security, but one needs encrypted digital storage to keep bitcoins safe. Storing bitcoins safely may be easier or cheaper than keeping cash safe at home, but it is likely more complex than relying on credit cards and bank deposits. Banks or payment services providers tend to be more reliable than storing bitcoins, due to their experience, well-developed systems, and the insurance they offer directly or indirectly. Of course, as the Bitcoin system matures one may imagine development of more secure storage options and services. Bitcoin is still a young currency, and one could argue that banks were not particularly safe early in their history due to theft from the outside and fraud from the inside.

Finally, unlike cash, Bitcoin cannot be counterfeited, so if you get it in a transaction, you can rest easy that it is genuine.[29] The bitcoin may be stolen, but the transactions are not reversible (unlike credit cards), so this is not a concern for the seller. Moreover, as Bitcoin is entirely managed by a well-defined algorithm it cannot be manipulated or tampered with by governments or other entities. Not all countries have put in place safeguard measures to avoid government interference in the management of the national currency, like central bank independence. The impossibility by any third party to alter transactions or manipulate the Bitcoin system could prove helpful for countries that do not have a reliable banking system.

Thus, on some dimensions it is not clear whether cryptocurrencies have more convenient attributes than the older currencies. Whether it is easier to carry and transfer, or safer to store may depend on the preferences of the users and complementing infrastructure. But on others, they provide a clear improvement, like divisibility, durability, or risk or fraud and counterfeiting. Those attributes could make cryptocurrencies more useful for some uses, like micro-payments or remote international payments than older alternatives. But the benefit needs to be large enough for people to adopt it and use it aside (or instead) of the traditional banking system and credit card system.

[29] People sometimes argue that Bitcoin is risky because relying on it means putting your trust in an anonymous programmer (or programmers) whose true intentions are unknown. However, Bitcoin requires the well-known and well-understood cryptographic tools that also underlie much of traditional payment infrastructure, e-commerce, etc. This means that if we trust the encryption of online banking or retail (as most people do), we should have the same trust in bitcoin.

CHAPTER 5

The Rich Landscape of Crypto

Regardless of its ingenuity, Bitcoin is not exempt of some unpleasant externalities (e.g., the high electricity usage), drawbacks that may affect its economic viability (e.g., the deflationary pressure), or limitations due to some technical details (e.g., a low throughput). In other words, Bitcoin's design left room for improvements in many directions. What certainly had an important impact is that the Bitcoin code is available freely under an MIT license, which allows anyone to copy, modify and even commercialize the software.[1] With such freedom, it did not take a long time before a number of alternative cryptocurrencies (often referred to as "altcoins") appeared, trying to fix the real, and sometimes only perceived, weaknesses in the Bitcoin design. Other altcoins, taking advantage of the new technology, took it into new directions aiming at different functionality (e.g., decentralized file-sharing system).

As Bitcoin attracted attention outside of the cryptography community in late 2013, the number of altcoins skyrocketed. Some of these new cryptocurrencies are little more than a copy of Bitcoin (for example, Terracoin). Others differ in technical detail (e.g., Litecoin uses a different hashing algorithm than Bitcoin, and adds blocks more frequently, but is otherwise very similar). Yet others proposed more radical changes to the design, with the potential to meaningfully change the economics of

[1] https://github.com/bitcoin.

H. Halaburda et al., *Beyond Bitcoin*,
https://doi.org/10.1007/978-3-030-88931-9_5

the cryptocurrency (e.g., Zcash). While several thousands of altcoins have been created, most of them fail into obscurity, with few achieving success.

There are various ways to assess the success of a cryptocurrency: its price (or its market capitalization), the traded volume and the level of activity on its blockchain and presence on exchanges. The price of a cryptocurrency is obviously the first metric most people think of, and the first assessment that can be made is that cryptocurrencies are highly volatile. Unlike most financial assets like stocks, bonds or other instruments, it is not surprising for a coin to have its price changing by more than 10% in a single day. With such high volatility, cryptocurrency's price is not a reliable measure of the coin's success.

Both the traded volume and the activity level on the blockchain (i.e., the number of transactions) give a measure of the usage of the cryptocurrency, but they capture different things. Many, but not all, transactions on the blockchain indicate the use of a coin for transaction purposes, that is, use of the coin as a medium of exchange. But there are also many transactions to and from exchanges, which are the main gateway to buy and sell cryptocurrencies. Those transactions reflect value transfers, not usage of a payment system. In contrast, traded volumes are mostly reflecting the level of speculation and investment of cryptocurrencies and, except for withdrawals and deposits, transactions that take place in an exchange are not recorded on the blockchain but on the exchange's ledger.

Overall, while there has been a proliferation of cryptocurrencies with a wide range of attributes and with some cryptocurrencies being technologically superior to Bitcoin, this latter still appears as the most successful. Bitcoin's total market capitalization as of July 2021 is about $600 billion, more than twice that of the next cryptocurrency, ether. But such high levels of market capitalization also signal a shift in the way people look at those new coins. Rather than being seen as payment systems, cryptocurrencies are increasingly used as investment assets.

Speculation around cryptocurrencies undermined their capacity to become payment systems. High volatility, a frequent consequence of intense speculation, moved cryptocurrencies further away from the price stability needed by any payment system. Yet, the quest of creating a new payment system based on crypto is still an important driving force for new innovations in the crypto space. Interestingly, the goal of improving cryptocurrencies' appeal as means of payment was also the motivation behind the earliest improvements made to Bitcoin. Looking at the 2010s, the history of cryptocurrencies looks like a trial-and-error process, with a series of endeavors to either improve Bitcoin or help cryptocurrencies occupy a greater place in our lives.

5.1 Improving Currency Functionality

With trades being processed every 10 minutes at best, it was quickly observed that Bitcoin is not very appealing for small quick trades like buying coffee. In October 2011, Charles Lee created Litecoin intended as "the silver to Bitcoin's gold."[2] It was meant to be used for low value, quick transactions. Specifically, it was not meant to compete with Bitcoin, but rather work together to grow the appeal of cryptocurrencies for ordinary payments.

Litecoin achieves quicker validation of transactions by adding blocks to the ledger four times faster than Bitcoin, i.e., every 2.5 minutes. It increases both the speed of individual transactions, and throughput of the blockchain in general. However, shorter intervals between blocks increase probability of accidental forks, as it is more likely that two different miners would broadcast their blocks before the other block propagates throughout the network. Such fork results in two competing versions of the ledger, one of which would be later orphaned. This property also increases propensity to malicious attacks. Yet, with the cryptocurrency's focus on small-value transactions it may not be worth it for the attacker to bear the cost of attack. The users may also be willing to take more risk of transaction failing with small values.

Litecoin also aimed to improve Bitcoin's problems of the "arms race" among the miners' excessive energy use and mining concentration by utilizing a different hashing algorithm for the proof-of-work—*scrypt* instead of SHA-256, used by Bitcoin. Scrypt requires relatively less computing power, lowering the amount of electrical energy needed to calculate the same number of hashes. Thus, it was possible to mine Litecoin using standard PCs at a time when mining Bitcoin competitively already required specialized equipment.

Litecoin's alteration was well-meant, but did not change the incentives of the participants in the cryptocurrency ecosystem and ultimately failed to resolve the arms race problem in mining. The underlying algorithm still has the tournament structure that rewards the miner with the most powerful machine, at least on average. As Litecoin's price increased, and with it the mining reward, miners had a stronger incentive to invest in more powerful mining equipment. Soon, ASIC equipment specialized for scrypt hashing function appeared in the market. Nowadays, it is virtually

[2] McMillan (2013).

impossible to mine Litecoin with a PC, because it is primarily mined by ASIC miners.

Litecoin also offers a larger supply of coins. While the supply of bitcoins is limited to 21 million, there will be many more litecoins created—84 million in total. This change was proposed to address the deflationary pressure present in Bitcoin's system. Unfortunately, increasing the total number of coins four times relative to Bitcoin does little to change the deflationary incentives. The supply is still finite and will stop growing at a known date.

Overall, while Litecoin recognized several important flaws in the design of Bitcoin, the only real improvement was increasing throughput and transaction speed. Since both Bitcoin and Litecoin have the same block size with the same maximal number of about 4000 transactions per block, Litecoin's maximal throughput is then 28 transactions per second. Yet, waiting 2.5 minutes for a small value transaction, like buying coffee, is arguably still too long. Feathercoin, a cryptocurrency introduced by Peter Bushnell in April 2013, added a block to the blockchain every minute. Otherwise, it had a very similar design to Litecoin. The total supply of coins was four times larger than Litecoin's, i.e., 16 times larger than Bitcoin's. It also introduced a new hashing function, Neoscrypt, to democratize mining and protect against the arms race, excessive energy consumption and mining concentration. Unfortunately, as we saw with Litecoin, these design choices are not an effective way to undo the deflationary pressure or the tournament structure inherent in the design of the currency.

Nonetheless, even though some of the attempted improvements failed, on throughput and speed of transaction Feathercoin was better than Litecoin, and both of them were better than Bitcoin. Yet, Bitcoin is the most successful cryptocurrency. Litecoin is still quite active, but Feathercoin is for all practical purposes defunct. Success of a cryptocurrency is difficult to predict—as we had seen in the discussion on competition between money in Chapter 2, network effects and excess inertia may play as big of a role in this dynamic as improvement in quality.

Another conclusion that emerges from the history of Litecoin and Feathercoin is that the development of alternative hashing functions does little to alleviate the arms race, excessive energy consumption and concentration. Even at the introduction of Neoscrypt with Feathercoin, it was admitted that it will not solve the problem of ASIC completely, it may merely postpone it into the future. This should not be too surprising.

Any proof-of-work algorithm will favor higher computational power and will give an advantage to whoever wields such power, no matter how small the difference between that miner and the next more powerful miner. This gives miners incentives to engage in an arms race, which in turn may give an incentive for hardware producers to develop mining rigs specialized for Neoscrypt as soon as they see enough demand.

The problems of the arms race, excessive energy consumption and mining concentration arise because of the tournament nature of the proof-of-work system in the early cryptocurrencies. Once it became clear that neither Litecoin nor Feathercoin could solve this issue, some subsequent cryptocurrencies experimented with stepping away from the proof-of-work system. To do this successfully, they needed to come up with a setup that would automatically check that proposed transactions are valid (based on the information from the previous transactions) and add them to the ledger.

The list of alternatives to proof-of-work is constantly growing, and one can get easily lost between all those concepts with similar names: proof-of-stake, delegated-proof-of-stake, proof-of-burn, proof-of-reputation, proof-of-authority, proof-of-elapsed-time, proof-of-time, etc.[3] Among all those alternatives, proof-of-stake (PoS) is one of the oldest and the one that has received the most attention. There are now a number of cryptocurrencies that are using PoS, like Peercoin or Tezos, or are planning to adopt it in the future, like Ethereum or Cardano.

5.2 Proof-of-Stake

While proof-of-work awards the first miner to find a valid nonce, proof-of-stake instead distributes the reward to all holders of a cryptocurrency, with people who hold more coins (i.e., those who have more stake in the system) receiving a larger "dividend." The first cryptocurrency using proof-of-stake was Peercoin, established in August 2012. But Peercoin was using proof-of-stake for some blocks, and proof-of-work for others. Nxt, established in November 2013, was the first cryptocurrency solely based on proof-of-stake.

The basic principle of PoS is relatively straightforward. At any time a miner (also called a validator or a minter) is randomly selected to be the

[3] https://medium.com/hackernoon/consensuspedia-an-encyclopedia-of-29-consensus-algorithms-e9c4b4b7d08f.

one building and adding the next block of transactions. To be selected, a validator must "freeze" a certain amount of coins, called the *stake*, in a staking wallet. The probability to be selected is proportional to the total stake. For instance, if Alice, Bob and Carol's stakes are 500, 300 and 200 coins, respectively, then Alice will be selected as the validator 50% of the time and Bob and Carol 30% and 20% of the time, respectively. Each time a validator validates a block he or she gets a block reward, consisting of transaction fees and newly created coins.

Similarly to proof-of-work, proof-of-stake provides security for the system by making it costly to counterfeit transactions by rewriting the blockchain. However, proof-of-stake does not follow Nakamoto's principle of one CPU, one vote. But in effect, there might not be much difference with Bitcoin's proof-of-work; the probability to be the miner adding the next block is also increasing with respect to the miner's "wealth." In Bitcoin, it is the miner's capacity to invest in mining rigs and energy cost—in proof-of-stake cryptocurrencies, it is the validators capacity to acquire more coins to increase his or her stake. And similarly to Bitcoin's 51% attack, an entity that controls the majority of a proof-of-stake currency will be selected more often than anybody else and could potentially modify the blockchain. In both cases, a high price of the cryptocurrency increases the cost of such a majority attack. While the former requires substantial computations that would be difficult to reproduce for an attacker, the latter requires establishing a large holding of a cryptocurrency, which would not only be costly for an attacker to gather but would only align incentives with the rest of the system (undermining a currency is less attractive when one is holding a large stake in that currency).

The proof-of-stake methods used in Peercoin and Nxt go a long way to solve the negative side effects of proof-of-work systems: excessive energy consumption and the mining arms race. The reason for that goes back to the economics of these systems and the incentives they create. As we saw earlier, the proof-of-work externalities have to do with the tournament structure of that system. In contrast, proof-of-stake does away with the tournament and selects the winner randomly, based on the number of coins they hold. The fact that there is only one validator at a time solves a great part of the consensus problem. With only one chosen validator the risk of having two proposals at the same time (which would lead to

a fork of the blockchain) is greatly diminished.[4] That allows for higher throughput of transactions, as larger blocks no longer mean there will be more accidental forks and orphaned transactions. Moreover, since validators thus compete in terms of stakes and not in terms of energy spent, high energy consumption is thus no longer necessary (the only energy spent is used to check validators' block proposals, which is computationally trivial). This means that the winner has little incentive to invest in state-of-the-art computing systems, solving both arms race and excessive energy consumption problems.

Proof-of-stake does not change the issue of deflationary pressure. Whether a system suffers from it or not, depends on the schedule of new coin creations in block rewards. Peercoin's algorithm set up a steady increase in the total number of peercoins at 1% per year (in contrast to the limited number of bitcoins). At the same time Nxt keeps the money supply static, with all coins being pre-mined and allocated across the initial users of the system. This means that all Nxt transactions must be accompanied with fees, which are then earned by the network nodes validating these transactions based on their proof-of-stake.

Overall, it seems that the proof-of-stake innovation is a clever solution to the excessive energy consumption and mining arms race experienced by Bitcoin and other "PoW" (proof-of-work) cryptocurrencies, and may also help alleviate some other problems, like limited throughput. But as it is often the case, with new solutions come new challenges. There are several major issues with the proof-of-stake design that were not present in proof-of-work.

The first and perhaps the most famous issue is the so-called *nothing at stake* problem. As we have seen earlier in the chapter, a key objective in Bitcoin's design is to ensure consensus, that having all miners relying on the same version of the blockchain. Proof-of-work is the first part of the design to ensure consensus. Since mining is costly, miners have an incentive to mine on only one branch of the blockchain (and the longest

[4] The risk is not zero, though. Each time a block is added to the blockchain the next validator has a certain amount of time to issue the next block, after which a new validator is chosen (in general, for each block, a sequence of validators is chosen that will be used if the first validator fails to propose a new block within a certain time frame). Lack of consensus can occur when, for instance, due to network latencies, the block mined by the first validator arrives shortly after the allowed delay.

chain rule ensures that all miners choose the same version). In proof-of-stake, in contrast, the stakes validators commit to are not lost if multiple versions of the blockchain coexist. Thus, validators have no incentive to focus on only one version, if the blockchain ever forks. Instead, they can validate a block when they are selected, whatever branch it is on. And thus there is no reason why any version of the blockchain would ever be orphaned.

Yet, while the validators do not lose anything in terms of the cryptocurrency coins by following this strategy, they may be aware of the negative externality that multiple versions of the blockchain imposes by decreasing the value of the cryptocurrency.[5] In other words, maintaining different versions of the blockchain may decrease the price of the coin, and thus reduce validators' revenue. This threat of lost value is a sufficient deterrent if the system restricts the rotation of validators to those with sufficiently large stakes, and if the block reward is not too high. Imposing a large minimal stake for validators ensures that the cost of lack of consensus is sufficiently high. A low block reward plays a similar role as it decreases the gains from maintaining disagreement.

A second issue with proof-of-stake is *grinding*, a tactic that validators may employ to ensure to be (nearly) always selected as a validator. As we have explained earlier, the basic principle behind PoS (proof-of-stake) is to select validators randomly (with the probability depending on validators' stakes). A key problem with PoS is to communicate to all the validators the "identity" of the validator in charge of the next block, ensuring that all validators agree on this identity. The exact description of the grinding strategy depends on the exact details of how the PoS mechanism is built, but usually the "identity" of the validator for the next block is determined using a function that depends on the hash of the current block, so that the identity of the next validator (his/her public key) is known to all. Since the hash function is unpredictable, the choice of the next validator looks random. Grinding consists of performing some changes in the block to be proposed until the selected next validator is a desired one (e.g., the same or a different wallet of the same validator). In the extreme case, grinding could allow a validator to take control of the blockchain forever. Peercoin and Nxt were both using a PoS algorithm that makes grinding feasible.

[5] Saleh (2021).

A solution to this problem consists of limiting validators' capacity to influence the outcome of the hash, that is, the source of randomness. So far there is no perfect solution to that problem, but several proposals substantially mitigate it. The first one consists of requiring validators to commit well in advance. That is, the random mechanism used to determine the next validator uses information that is found in blocks added some time ago, say, 20 blocks in the past. Under this solution a validator can ensure to remain the validator for the 20th block in the future but not for the next 19 blocks. Another solution, quite more technical, consists of using advanced cryptographic tools (e.g., differential cryptanalysis) to detect deviations from the expected probabilities of miners being selected. The drawback is that one detects grinding ex post rather than preventing it.

A third problem with PoS is the *long range attack*, which happens when a validator forks the blockchain by going way back in the blockchain's history (and then uses a grinding attack to construct a blockchain longer than the legitimate one). Such an attack can be possible when former validators sell their, by now empty wallets (or past credentials to be a validator) to an attacker. The attacker then can go back in the history of the blockchain to a block for which one of those former validators was the designated validator and fork the blockchain starting from that block. The attacker will be able to fork for two reasons. First, by being now in possession of the private key of the former validator the attacker will be able to sign the block as if the attacker were the true designed validator. Second, a major difference with PoW is that under PoS it takes virtually no time to make a new valid block. There is thus no uncertainty about whether an attacker will be able to build an alternate version of the blockchain that is at least as long as the legitimate version.

Since deviations are virtually costless under a PoS, the only way to incentivize validators to follow the protocol is to introduce a penalty that can be imposed to wrongdoers. There are several challenges with penalties, though. First, deviators must be identified and there must be some mechanism in place to enforce the penalty imposed to deviators. While it is relatively easy to identify the validators behind each block, it is more difficult to identify validators that accept fraudulent blocks or keep a copy of a fork and thus help the dissemination of the fork. Second, penalties must be enforceable. In principle this is where the stakes deposited by validators can help: it suffices to liquidate the wrongdoers' stakes. There are several issues with this method, known as "slashing." For one, the

loss of the stake may be lower than the gain from the fork (e.g., the benefit from double spending), making any threat of losing the stake moot. Another issue is that one needs to guarantee that the loss of the stakes is done not only in the main, legit version of the blockchain but also in the version built by the attackers. A third challenge is that deviations need to be observed, preferably as quickly as possible, so that an alternate version of the blockchain does not become de facto permanent.

Recently new types of proof-of-stake algorithms are developed, like Cardano and Ethereum 2.0, which promise reliable solutions to these main incentive issues. It is yet to be seen how these systems perform on a large scale in real life.

5.3 Privacy Coins

Cryptocurrencies discussed earlier in this chapter improve upon Bitcoin in order to make cryptocurrencies a more appealing and affordable means of payment. But the crypto landscape has also seen many cryptocurrencies developed for specific purposes, like increasing privacy, promoting tipping, facilitating peer-to-peer file sharing, or enabling more complicated operations via smart contracts. The category seeing the most innovation are privacy coins.

Bitcoin is often thought to be the digital equivalent of cash: anonymous and hardly possible to trace once spent. This is at best a simplification. As we discussed in the previous chapter, Bitcoin's blockchain offers an exact and transparent record of the Bitcoin transactions, which means that Bitcoin is more correctly described as a "pseudonymous" rather than "anonymous" currency. In practice, few observers would be determined enough, or would have enough resources, to track the transactions and the Bitcoin holdings directly to the real-life people involved. This makes the currency opaque enough to be anonymous enough even for some nefarious purposes. Nonetheless, we had seen examples where institutions with enough resources tracked the movement of bitcoins closely enough to identify individuals controlling the wallets. This is how the FBI identified Ross William Ulbricht as the head of Silk Road, for which he was sentenced to life in prison. Blockchain transaction records also facilitated the sentencing of several Silk Road merchants of illegal substances.

Consequently, multiple cryptocurrencies sought to improve on protecting the users' privacy and increasing the anonymity of transactions. Two of the first cryptocurrencies aimed to provide a higher degree

of anonymity than Bitcoin were Darkcoin (rebranded Dash in March 2015) and Cloakcoin. Darkcoin was introduced as XCoin in January 2014, changed the name in February 2014 to Darkcoin, and in March 2015 it changed name to Dash. It increases anonymity of transactions by bundling transactions through a process called coin mixing. For example, instead of two separate transactions, A to B and X to Y, the ledger reflects only one transaction A&X to B&Y, obfuscating the individual transaction links. The problem with straightforward coin mixing is that the transaction inputs and outputs can be matched by a size. If A sends 2 Dash and X sends 5 Dash, while B receives 2 Dash and Y receives 5 Dash, the transactions can be matched, even with coin mixing. Darkcoin countered it with pre-mixing denominations already in the wallet and combining identical inputs, so that inputs cannot be matched to outputs. For example, A's wallet can send two independent transactions (and thus cannot be connected to the same sender) and X's wallet sends five independent transactions, 1 dash each. Then there are seven independent 1 dash transactions, each received by a separate address, so it cannot be seen directly that B received 2 dashes and X received 5 dashes.

Cloakcoin, introduced in May 2014, also uses mixing but to a higher level than Darkcoin. To ensure anonymity, each transaction uses a unique stealth address and is mixed with other transactions provided by other nodes (called "cloakers"). All those transactions are merged together into a single transaction which is sent to the network to be added to the blockchain. Over the years, Cloakcoin has been substantially redesigned. Cloakcoin's main website is still updated and is mentioning updates planned for 2021 and beyond, suggesting that the project is still ongoing. Yet, it seems that the coin has not attained the success its developers hoped. Both the price and the trading activity of Cloakcoin are very low.

Mixing is an anonymizing strategy that is not confined to coins that are designed explicitly for that like Dash or Cloakcoin. There are now a number of companies offering mixing (also called tumbling) services for Bitcoin. Those services break up a transaction into several transactions, merge them with other transactions and transfer them through several transactions to the intended recipient. Although such mixing is less sophisticated than Dash or Cloakcoin, it can still be relatively effective and make it nearly impossible to trace with precision transactions on the blockchain.

Two other anonymity coins, Zcash and Monero are part of a new generation of cryptocurrencies that pioneered the use of advanced cryptographic tools to obfuscate transactions—which were developed independently of Bitcoin. Zcash, built by Matthew Green and Zooko Wilcox, based on Bitcoin's code, was released in October 2016. Under Zcash users can choose to have their transactions either transparent (like in Bitcoin) or "shielded." In shielded transactions only the fee amount is public, but the sender and receiver's address, as well as the amount of the transaction are obfuscated. To achieve anonymity of transaction, Zcash relies on a zero-knowledge proof protocol, which is a protocol that allows someone to prove knowledge that some statement is true without revealing it. Recall that in Bitcoin (or for a transparent Zcash transaction) miners validate the transactions before adding it to a block. That is, miners will check that the sender has sufficient funds and that he or she is indeed the owner of the wallets from which the coins are sent. Such verifications are done using the senders' public key, which is observable by the miners. A zero-knowledge proof protocol like the one used by Zcash permits the sender to prove to the miners that a shielded transaction is valid even if the miners cannot observe it. Zcash's anonymization protocol is relatively rich as it allows transactions to mix shielded and transparent inputs and outputs. For instance, a transaction can have a mix of shielded and transparent inputs, in which case only the amount corresponding to the transparent inputs will be visible. Another feature is that users can opt for a selective disclosure of their shielded transaction for auditing purposes.

Monero is another cryptocurrency that provides anonymity of transactions. Its design was first laid out in a white paper written in October 2013 by Nicolas van Saberhagen (a pseudonym whose real identity is unknown). Monero went live on April 18, 2014. The basic design of Monero (which means *coin* in Esperanto) is similar to Zcash, where validation of obfuscated transactions is done with a zero-knowledge proof protocol. But Monero goes further first by obfuscating all transactions and also by controlling how transactions are broadcast to the network so as to eliminate the possibility to trace a transaction (e.g., through the senders' IP address).

Although Monero and Zcash are similar in the way transactions are anonymized, they adhere to different notions of privacy. Contrary to Monero, Zcash is compliant with US and European anti-money-laundering policies, which includes, for instance, customer due diligence, reporting suspicious transactions or providing required originator and

beneficiary information for virtual asset transfers by virtual asset service providers (e.g., exchanges). Complying with such regulations may seem at odds with the design of an *anonymous* cryptocurrency, but it is not. The objective under Zcash's design is to guarantee users a privacy level similar to that of, say, cash, but not to evade law enforcement.[6] One can thus interpret Zcash as an attempt to offer a digital analog—yet decentralized—of fiat currencies while Monero is aimed at offering a level of privacy that is not provided by fiat currencies.

The level of privacy offered by a cryptocurrency goes beyond a mere adherence to some moral principles; it generates a tradeoff that can impact the evolution of the coin. On the one side of the tradeoff, compliance with regulations may make privacy protection more difficult to guarantee. One can thus legitimately argue that this runs contrary to the original principles of cryptocurrencies which aimed to protect individuals against governments' possible control over financial transactions. This is, for instance, why the high level of anonymity made Monero increasingly more attractive in darknet marketplaces or ransom attacks.

On the other side of the tradeoff, this higher usage can attract regulators' attention and increase the risk of restrictions. Indeed, regulatory pressure led several exchanges to delist anonymous cryptocurrencies. Cutting off links between crypto and fiat currencies makes it more difficult to ultimately cash out fiat money, thereby undermining the attractiveness for such cryptocurrencies. So far regulators and law enforcement agencies have not been able to fight effectively against anonymous coins but the pressure is likely to intensify as such coins become more popular.[7]

5.4 Proliferation and Eventual Decline of Altcoins

In our short review we have discussed a number of Bitcoin's cousins. This, however, merely scratches the surface. Many hundreds of cryptocurrencies that are basically copies or clones of Bitcoin, Litecoin or Peercoin have been created. For example, Zetacoin and Monacoin are based on Bitcoin; Infinitecoin, Goldcoin, and Ekrona use Litecoin's design, etc. Most of

[6] https://bitzec.github.io/wp-content/uploads/2019/04/Zcash-Regulatory-Brief.pdf.

[7] See Greenberg (2017) and Murphy (2021).

these cryptocurrencies use the same technology and do not offer their users any meaningful improvement over the earlier ones, and hence are called *copycat currencies*.

Most of such cryptocurrencies were introduced following a sharp and spectacular increase of Bitcoin's price in late 2013. The trend lasted until 2017. Creating a new cryptocurrency is relatively easy. Bitcoin being open source, anyone can reuse the same algorithm and code to create a similar cryptocurrency, driving the cost of creating a new cryptocurrency to very low levels.

By 2014 websites even sprang up that would, for a fee, help you set up a new altcoin automatically, which further decreased the cost of creating altcoins. An example is Coingen.io—now defunct—that allowed users to automatically generate an altcoin based on Bitcoin by choosing the key variables (e.g., how often a block is added to the blockchain, how many coins successful miners get, how quickly the reward for mining decreases).

While the ease of entry may explain the multitudes of altcoins offered to the market, it is more difficult to see why those new cryptocurrencies would get any traction given that most were not offering any substantial improvement. One reason could be that miners were looking for alternatives. They may be discouraged from participating in the older schemes because they lack the specialized gear (or funds to buy such equipment) necessary to have a chance to be successful when mining bitcoins or litecoins. Instead, such miners may be looking for newer, less crowded coins to mine, because they stand a higher chance of successful earnings with such currencies. They might then hope to sell such cryptocurrencies in digital exchanges.

Of course, the above argument only begs the question why anybody would buy the copycat coins from such miners. It is possible that some people may trade them as an experiment, perhaps to get to know the industry, and they perceive such cryptocurrencies to be more accessible than, say, Bitcoin. They may also see them as potential contenders to Bitcoin, betting on the "next big thing." That may explain why prices of all cryptocurrencies increased during Bitcoin's price increase in 2013.[8]

It has been suggested that prices of some of these cryptocurrencies may actually be driven by pump-and-dump schemes. Cryptocurrencies typically start with a number of coins that are already pre-mined. That

[8] See Gandal and Halaburda (2016).

is, these coins are created before the first block on the blockchain, and before the cryptocurrency is brought into the mining community. Later, as the miners mine new coins and sell them on the market, the owner of pre-mined coins buys a lot of them, to increase the price. (That's the "pump.") As the price increases, the altcoin attracts attention. As more people see it as a potential success, they might want to participate in it or even start viewing it as an investment. When they buy some units of the cryptocurrency, they typically buy them from the creators of the scheme, who choose this opportunity to cash out. (That's the 'dump.'). Afterward, the price usually drops, and never recovers.[9]

Most of the altcoins have become defunct over time. A precise count of defunct cryptocurrencies is difficult to have, for the simple reason that many cryptocurrencies became defunct shortly after their launch, having been barely noticed. However, according to various estimates it is believed that there are several thousands of defunct cryptocurrencies.[10] In fact, the fates of early cryptocurrencies seemed to follow a pattern similar to many websites and startups created during the dotcom bubble in the early 2000s. That is, while Bitcoin reached unprecedented high prices and Litecoin maintained a robust presence, many ventures have either disappeared or became negligible according to the coin's price (or market capitalization), the traded volume or the activity level on the blockchain. Moreover, it is difficult to find any significant correlation between popularity during the early years and long-term success. Nxt became a niche currency, with its price dropping to a penny. Peercoin, Feathercoin, and Cloakcoin do not show any real activity. The blockchains of these three coins are still constantly updated, but apart from the coinbase transaction generating new coins that accrue to the miner, virtually no other transactions are made and there is minimal activity on exchanges regarding those coins.[11] Deducting that those coins are about to disappear would be a hasty conclusion, though. For instance, Cloakcoin has updates scheduled for 2021, which suggests that the developers behind it still consider it as an ongoing project. Some of them may also experience a comeback similar to Dogecoin's.

[9] See Gandal et al. (2018).

[10] See https://99bitcoins.com/deadcoins/ or https://www.coinopsy.com/dead-coins/.

[11] Feathercoin's blockchain can be explored here: https://explorer.feathercoin.com.

Dogecoin was created by Billy Markus and Jackson Palmer in December 2013. At the time, Bitcoin had gained substantial popularity and presence in the media, and notoriety—coming from sensational origins from unknown Satoshi Nakamoto and use for illegal activity as became apparent with the Silk Road bust in the fall of 2013. This newly gained notoriety made Bitcoin an interesting concept to read or hear about, but possibly not an innovation you would want to be a part of. Markus and Palmer wanted to change that, and thought of a cryptocurrency design that would be "cute" and more "fun to use." To make their new currency more fun, they associated it with an image of Shiba Inu dog; the name of the currency is also derived from a misspelled, or perhaps spelled in a cooler manner, word "dog."

The cryptocurrency was initially proposed as a "tipping coin": available in large quantities, with a relatively low price per unit. The goal was to make it suitable for philanthropy, charity, and tipping—in essence, the equivalent of a "Like" or "+1" button that would convey a small monetary reward. There were several other such tipping coins created around the same time, like Karmacoin or Reddcoin.[12]

Dogecoin was designed with its intended use in mind. Its original algorithm was borrowed from Luckycoin, a "casino currency" that randomized the mining rewards, presumably to make using the currency more exciting for its users.[13] However, because this feature created uncertainty about the cost and benefit of mining, it did not catch on in the Dogecoin community. Consequently, in February 2014 the rewards for mining were set to the fixed limit of 250,000. The total number of dogecoins to be created was initially thought to be fixed at a relatively large number (100 billion), promising enough units of the currency to support

[12] Of course, a similar tipping method could be designed using Bitcoin, given its divisibility. However, the concern would have been that the negative aspects of bitcoin's reputation would make people less likely to use the currency in this manner ("to do good and to feel good"). Moreover, the denomination matters psychologically: sending or receiving 100 dogecoins may well feel better than, say, 0.00006 bitcoins, even if the value of the gift is the same in the units of the state currency (say, the dollar).

[13] Luckycoin was a modification of Litecoin with the added feature that randomized the reward for mining a block. The standard reward for each block is 88 luckycoins. However, with a 5% probability the miner could get twice as many, with a 1% probability five times, and with a 0.01% probability, 58 times as many coins as the reward for mining a new block.

tipping, etc. Due to a quirk (likely a mistake) in dogecoin's programming, however, the algorithm was set to keep awarding a fixed number of dogecoins per block indefinitely, making the supply of the currency increasing over time and potentially unbounded. The Dogecoin community has decided not to remove this feature. These changes to the design of Dogecoin illustrate the amount of experimentation that was involved in creating cryptocurrencies, as well as potential for unintended mistakes. We can also see from Dogecoin's example that if the blockchain is maintained by a smaller community, changes are possible to implement without forks.

Dogecoin or other tipping coins did not get adopted as intended. As many other cryptocurrencies, Dogecoin declined in price and activity after 2016, and lingered in relative obscurity. Yet, unlike most other cryptocurrencies, Dogecoin turned out to be much more resilient re-emerging in early 2021. At that time, it became a target for a pump scheme by several followers of the Reddit group r/WallStreetBets.[14] Several tweets by Elon Musk promoting the coin followed suit, pushing Dogecoin among the highest cryptocurrencies in terms of market capitalization. While public statements made by Elon Musk in 2021 have largely contributed to the prominent place that this coin reached in the cryptosphere, it is difficult to pinpoint the exact reasons why Dogecoin interested Musk.

5.5 The Emergence of Tokens

The number of cryptocurrencies exploded after 2013. But most of them declined and even disappeared by 2017. Yet, the list of traded coins on crypto exchanges is longer than ever. This is because a large number of traded coins are crypto-tokens, not cryptocurrencies.

Crypto-tokens—typically called just *tokens*—emerged en masse in 2017 on Ethereum, a blockchain platform supporting a wide range of functionalities through smart contracts (we discuss technical aspects of tokens and smart contracts in more detail in the next chapter). A smart contract on Ethereum allows to keep track of tokens transaction ledger and prevent double spending without the need to run a separate blockchain for the coin. This makes launching a token a much simpler task than launching a cryptocurrency, as there is no need to attract miners to a new, separate blockchain. Only cryptocurrencies with high value attract a sufficient

[14] This Reddit group is the same group that launched the rally behind Gamestop's shares in 2021.

number of miners to make the blockchain secure. In contrast, a token on the Ethereum blockchain automatically inherits the security provided by that blockchain.[15] The Ethereum miners still need to be paid to maintain the token's ledger, but that payment is shared with other activities on the same blockchain, e.g., other tokens.

While the cryptocurrencies and tokens differ on the technical structure, the differences are not noticeable when it comes to trading. Hence, they are indistinguishable on crypto exchanges. Out of the 5528 coins listed on coinmarketcap.com in July, 2021, 4435 were tokens and "only" 1093 were cryptocurrencies. This is also why they are jointly referred to as "crypto" or "coins."

The transition from cryptocurrencies to tokens coincides with a change in the objective behind the creation of coins. At the beginning a significant driving force of the competition between coins were the technical aspects related to use as general purpose currency (throughput, anonymity, mining protocol, etc.). As we had seen, with time, cryptocurrencies focused on special purpose use like anonymity, tipping or access to prespecified services. For example, Filecoin is a cryptocurrency aimed at facilitating sharing of hard drive space. Users storing files via Filecoin blockchain pay storage in filecoins, and those lending hard drive space are remunerated in filecoins.

This trend has been expanded with tokens, due to the ease of their deployment, integration with other software and programmability. The years 2017 and 2018 experienced an explosion of proposals on how the tokens can be used. The variety of activities, goals or services proposed in relation to the use of tokens or cryptocurrencies is large, ranging from transaction fees (Binance coin, to pay lower transaction fees on the Binance exchange), voting token for managing governance of decentralized apps (REP for Augur), fundraising for sustainable energy projects (Sustainable Energy Token), rewards for workout challenges (PUML Better Health), payment and reward system for family related services (Baby Token), and many more. The optimal design of coins' functionalities depends on the purpose they are expected to serve, especially if they are a part of a larger service platform. This is reminiscent of the design choices for platform-based digital currencies discussed in Chapter 3. Interestingly, tokens issued on Ethereum make it convenient to issue or sell the

[15] Although most tokens are issued on Ethereum, some tokens are issued via smart contracts on other blockchains, like Binance Chain, Neo, or Stellar among others.

token—in an event called Initial Coin Offering, ICO—before the related application or service is created.

For what it seems, the fate of cryptocurrencies and tokens are similar. Cryptocurrencies did not replace older payment systems in general, although they are the currencies of choice for some illegal activities like drug sales or ransom, and some niche legal markets like private aviation. The properties offered by cryptocurrencies did not seem to answer any important need or be superior in a meaningful way to already existing payment systems. Rather, cryptocurrencies are increasingly used as an investment vehicle. As for tokens, most of the platforms and services proposed during the ICO boom of 2017–2018 were not created at all, are still in development, or have gained very little traction. In effect, few tokens are used for their stated purpose. Instead, like cryptocurrencies, tokens are increasingly traded as an investment asset.

5.6 Stablecoins

Perhaps one of the most visible things about cryptocurrencies is their high volatility: daily drops and spikes of their exchange rates of more than 10% are not uncommon. Such variations are not only observed in the exchange rate between cryptocurrencies and fiat money but also between cryptocurrencies (e.g., the price of, say, Bitcoin in ether). Without any stable exchange rate, there is little hope to see crypto used as anything else than investment and speculation tool. Tokens, being governed by smart contracts and without the burden of maintaining their own blockchain, may offer a solution with *stablecoins*. Stablecoins are tokens that are aimed to offer, as their name suggests, a stable exchange rate between crypto and fiat money, similar to what some countries like Argentina did when pegging the peso to the US dollar in 1992.[16]

There are three main ways stablecoins aim to achieve stability. The first way to stabilize the exchange rate of a coin consists of increasing or decreasing the supply of coins so as to maintain the sought-after equilibrium price. This was the design of Basis, a (now defunct) stablecoin. A major challenge with this stabilizing method is that one needs deep pockets to prevent the price from falling below the target price in case

[16] The rate was 1 peso for 1 dollar and was maintained until 2001.

large amounts of the coin are being sold.[17] This need for large amounts of capital may be the reason why this mechanism is not commonly used in stablecoins.

The second, and more common stabilization design consists of using the fiat currency to which the coin is pegged to as collateral. Each coin then represents one unit of the collateralized fiat currency, which, in principle, can be redeemed at any moment by the buyers of the token. Hence, a fiat-backed coin can be seen as a receipt of deposit of the fiat currency. There are a number of such stablecoins: USD Tether, Stasis Euro, JPYCoin.[18]

The third way to stabilize the coin's exchange rate is to use other cryptocurrencies and tokens as collateral. Thus, the price of the coin is pegged to an asset (usually a fiat currency) distinct from the collateral. Since the exchange rate does not come naturally from the collateral, the price is maintained by a group of individuals or bots that sell or buy the token so as to maintain its price. An example of such a coin is Dai, pegged to the US dollar but using several Ethereum-based coins as collateral (i.e., ethers or certain tokens on Ethereum). An individual who wants to acquire some Dai tokens sends some ethers to a smart contract and gets some Dai tokens in exchange. The ethers sent to the smart contracts are held in escrow until the issued Dai tokens have been returned. Crypto-backed stablecoins need to keep more collateral than just the value of issued tokens, to account for the price volatility of the collateral asset with respect to the target price of the token.

It is tempting to argue that fiat-backed and crypto-backed coins are similar because both types of coins rely on collateralization and are pegged to a fiat currency. There is a tradeoff behind the choice between those two methods. On the one hand, a fiat-backed coin is more certain to guarantee a stable price. As we have intuited above, pegging the coin with the

[17] This can happen, for instance, when there is an attack on the coin. This stabilizing mechanism has been used by central banks to peg their currency to other currencies, like Swiss franc pegged to Euro in 2010's or British pound pegged to Germany's deutschmark in the early 1990's. Yet, this mechanism proved to be costly and vulnerable to currency price attacks. The most spectacular such attack was George Soros' attack on the British pound in 1992.

[18] In fact, there is no obligation to use a fiat currency. For instance, Digix Gold Token (DGX) is a stablecoin pegged to the price of gold—1 token equals 1 gram of gold, where for each purchase of the token an equivalent amount of gold is stored in a vault in Singapore.

fiat currency used as collateral naturally guarantees a constant price. Each coin is redeemable at any moment with the currency initially deposited. In contrast, crypto-backed coins cannot guarantee a constant price. In fact, crypto-backed coins like Dai are usually presented as being only *soft pegged*, i.e., without an exact, constant exchange rate. Crypto-backed stablecoins also need to keep more collateral than fiat-backed ones for the same value of tokens issued; and thus are more expensive to maintain. On the other side of the tradeoff, fiat-backed coins pose a problem of trust in the custodian. Contrary to crypto-backed coins, for fiat-backed coins the collateralization happens outside the blockchain. This thus requires a third party to manage the collateralized asset. Trust in the custodian is a non-issue for crypto-backed coins since the custody of the collateral is guaranteed by the code of the smart contract.

The history of Tether, a major stablecoin pegged to the US dollar, shows that deception or fraud is not a remote possibility. Tether was found not to be fully backed by dollar deposits as it was claimed by its issuer, Tether Limited, a subsidy of Bitfinex (a crypto exchange).[19] In spite of several failed promises of transparency and a price drop, Tether nevertheless managed to regain interest and maintain a nearly constant price of 1 US dollar for 1 tether. This surprising stability would suggest that a common belief (about the target price) could suffice to maintain a fixed exchange rate in lieu of collateralization.

So far stablecoins are essentially attractive for trading at exchanges, as using stablecoins permits buyers and sellers of cryptocurrencies to anchor their transactions to a riskless asset. Someone selling, say, some bitcoins can do it in exchange of some stablecoin instead of another volatile cryptocurrency. In other words, stablecoins are perhaps the only cryptocurrencies that are used essentially as a means of payment.

One stablecoin that has received much attention in the media is Diem, formerly known as Libra, a project proposed by Facebook. Diem is proposed to be a US dollar backed stable coin (versions of Diem stabilized against other fiat currencies are planned for the future as well). However, the Diem project differs from most cryptocurrencies in several aspects. First, unlike most other cryptocurrencies, Diem's blockchain is permissioned. Second, Diem does not only consist of a coin but it is also a payment system, making it similar to services like Venmo or Paypal.

[19] https://ag.ny.gov/press-release/2021/attorney-general-james-ends-virtual-currency-trading-platform-bitfinexs-illegal.

This last aspect is one of the main reasons that explain Diem/Libra's tempestuous history. Indeed, initially announced as being called Libra and backed by a number of well-established companies (e.g., Visa, Uber, eBay, etc.), the project rapidly raised concern from regulators of various countries, and from the general public worried by Facebook's reputation when it comes to privacy. After the initial announcement in June, 2019, the project was redesigned in response to regulatory concerns, and its name changed from Libra to Diem. However, the project is still in progress, with no launch date available yet.

5.7 Trading Crypto

So far, we discussed developments of different cryptocurrencies and tokens, implicitly assuming that people who want to use them already have them from some source. We also covered one such source for cryptocurrencies: mining. However, few potential users of, say, Bitcoin can reliably get that currency from mining. As we explained, mining has become ultracompetitive and requires substantial resources and expertise from anybody who wants to do it successfully. Similarly, one may receive bitcoins by accepting payment in that currency. However, few people would want to start a business solely to acquire bitcoins to then spend them on a different good. As for tokens, they cannot be mined. Like cryptocurrencies, they can be obtained by accepting payment. Tokens can also be obtained directly from the issuer when they are first released—either through purchase, or when they are distributed to users at no charge. For instance, COMP token is distributed to the most active users of Compound Finance platform. Note that whether the token can be bought from the issuer or needs to be "earned" through some activity is a design choice reminiscent of the choices platforms like Amazon, Second Life or War of Warcraft make about their digital currencies (as described in Chapter 3).

Perhaps the easiest and most common way to acquire cryptocurrencies or tokens is simply to buy them from other people. While there are technical differences between cryptocurrencies and tokens, there is no difference between them when it comes to buying and selling them. Such transactions are most of the time performed at exchanges like Coinbase, Binance or Huobi Global to name just a few, and are conceptually similar to the trade of financial assets like stocks. Trades can be made between a coin and a fiat currency (e.g., bitcoins bought and sold in US dollars)

or between two coins (e.g., bitcoins bought and sold in ethers). Without exchanges, large-scale flows between cryptocurrencies such as Bitcoin and other currencies (both crypto and traditional ones) would be difficult, which would be a major impediment for cryptocurrencies to play a role in the economy.

Before we move on to exchanges, a very direct way to acquire a cryptocurrency is to find a seller directly. Such meet ups were the oldest way for people to acquire bitcoins without having to become miners themselves. Usually, people interested in trading would coordinate over the Internet, using message boards, email, etc. They would then meet "in the real world" and transact: the buyer would provide the traditional currency, and the seller would initiate the Bitcoin transfer.

The above description is reminiscent of an early form of exchange: barter. The problem associated with barter, coincidence of wants, arises here as well. If you'd like to buy bitcoins, you first need to find someone willing to part with them, for the amount of traditional currency that both of you find acceptable. Of course, modern technology makes this problem much easier to solve than it has been historically, but it is nonetheless a friction. One of the themes in our book is that such frictions spur innovation and catalyze new, improved designs. This time is no different: Bitcoin ATMs have appeared in a few countries, allowing for an easy exchange of the traditional currency for bitcoins. Bitcoin ATMs (more often referred to as BTMs) were one of the early attempts to democratize access to Bitcoin. But over the years the popularity of BTMs has hardly increased, mostly because exchanges have proven to be an easy and convenient way to acquire and sell cryptocurrencies.

An online crypto exchange is a two-sided platform that connects buyers and sellers and allows them to trade their crypto holdings (cryptocurrencies or tokens). Conceptually, an exchange is similar to a traditional financial exchange. The resemblance extends to operations, too. Crypto exchanges operate similarly to stock markets, like Nasdaq, using a protocol known as the *continuous limit order book trading*. Under this protocol individuals submit orders to buy or sell that indicate the direction—sell or buy—and a quantity. The two most common type orders under this protocol are the market order and the limit order. A limit order to buy also indicates a *bid*, which is the maximum price the buyer is willing to pay. Similarly, a limit order to sell indicates an *ask*, which is the lowest price the seller is willing to accept. Market orders do not contain a bid or an ask: the agent submitting such an order will accept

the price given by the market. Under this protocol an order that has been submitted is matched against orders that have arrived earlier but could not have been filled, which show up in what is called the *book of orders*.

To see how this works suppose, for instance, that the book only contains one order, by Alice, to sell 1 bitcoin with an ask of, say, $30,000. In the financial terminology Alice is said to be a *liquidity provider*: the presence of her order in the book is a signal that the market is *liquid*, that is, that there are people who are willing to trade. Shortly after Alice submitted her order, Bob is submitting a limit order to buy 3 bitcoins at a maximum price of, say, $30,500 per bitcoin. Since Bob is willing to pay more than what Alice is asking, a trade is feasible. Bob is then called a *liquidity taker*. Only 1 bitcoin will be traded between Alice and Bob, and at a price of $30,000. That trades are made using the price proposed by the liquidity provider is simply mimicking what happens in most markets: transactions are usually made at the price that was announced first. Since Alice only offers 1 bitcoin, Bob's order cannot be fully executed. His order is then updated, becoming a limit order with the same bid, $30,500, but for only 2 bitcoins. That order cannot be filled—Alice was the only seller—so this order is now stored in the book, waiting to be matched against an order from another seller.

The existence of crypto exchanges is important for the competition across the various coins. Being listed on an exchange is one of the main tools used to increase the visibility of a cryptocurrency or a token. The prices at which they trade on the exchanges can be interpreted as the market's assessment of the relative importance and value of each cryptocurrency. But the benefit obtained from being listed also puts coins in a weak position as it gives more bargaining power to exchanges.

Exchanges are free to choose which coins to list, and except for the major coins they usually charge listing fees, which may, for instance, depend on the expected daily or weekly volume and on the dominance of the exchange. For large exchanges like Binance or Bittrex, listing fees can easily reach several hundreds of thousands of dollars if not above a million dollars, although exchanges usually deny charging listing fees. Payment to the exchange often consists of giving a certain amount of the new coins to the exchange, with the benefit of increasing liquidity provision when the coin starts being traded.

By analyzing frictions in operations of crypto exchanges we can gain insights into the functioning of the whole crypto ecosystem—not only in terms of the quality of the financial infrastructure (the exchanges

themselves) but we can also gauge how much attention people pay to the various coins. For example, in a well-functioning market, prices on the exchange should reflect all the relevant information available about the coins. It is actually fairly difficult to test how efficient markets are from that perspective. However, irrespective of whether the market is more or less efficient, it is generally agreed that it should not allow what economists call "arbitrage opportunities." Arbitrage is a type of trade that guarantees an investor instantaneous profit without any risk. An arbitrage consists of buying some asset for some price and being able to sell it at the same time for a higher price. In well-functioning markets, persistent arbitrage opportunities should arise very rarely (or never). To the extent they do arise, they are usually caused by market fragmentation, a particular friction in the way people trade, or perhaps are just a testimony that the market is relatively small and that its participants do not pay enough attention to what is happening in it.

The first example of an arbitrage opportunity is when a cryptocurrency can be bought and sold using different currencies or coins; for example, when bitcoins can be exchanged for US dollars and euros or litecoins. There is an arbitrage opportunity when buying bitcoins for $10,000 yields a different number of bitcoins than first buying litecoins for $10,000 then using those litecoins to buy bitcoins. Suppose, for instance, that one gets 0.3 bitcoin when buying bitcoins with 1 dollar but 0.32 bitcoin when buying with litecoins. One can thus obtain a sure profit by acquiring first bitcoins through litecoins and then cash them out in dollars: 0.32 bitcoins, obtained initially with $10,000 can be sold for $10{,}000 \times (0.32/0.3) = \$10{,}666.66$.

Another example of arbitrage opportunity is when the same pair of assets (e.g., dollars and bitcoins) is traded in two different exchanges. If the price is not the same in both exchanges, then there is again the opportunity to make an instantaneous profit for sure.

Arbitrage opportunities are relatively frequent in most financial markets, but usually do not last long.[20] Traders are constantly checking asset prices across several markets and rush as soon as arbitrage opportunities occur, which disappear when traders make use of them. In the above example, traders would rush to buy litecoins and then bitcoins, creating two converging dynamics. The first dynamic is an increase of the price

[20] See Budish et al. (2015).

of bitcoin in litecoins and the price of litecoin in dollars to increase, due to a higher demand. That is, one would get less than 0.32 bitcoins. The second dynamic is a decrease of the price of bitcoin in the bitcoin/dollar market due to a higher supply. That is, one would get less than $10,000 for 0.3 bitcoins. Those two dynamics stop when the arbitrage opportunity stops, i.e., when the number of bitcoins is the same whether bought in dollars or in litecoins.

Bitcoin and other coins can in principle function independently of traditional financial institutions. Early crypto exchanges, which involved only crypto-to-crypto trade, could also afford such independence. But if a crypto exchange offers fiat-to-crypto trade, it must have an account with a traditional bank to hold the fiat. Thus, crypto exchanges are now commonly linked to the traditional financial system, allowing users to fund their accounts with fiat currencies to then acquire crypto, or conversely, to sell their crypto and then withdraw the fiat currency from the exchange. Importantly, exchanges provide an environment in which trade takes place, but may not (and commonly do not) participate in these trades directly; exchanges are intermediaries that provide the service of matching buyers and sellers willing to transact at a given price.

Although crypto exchanges are in many ways similar to modern financial markets their infrastructures pale in comparison to exchanges like Nasdaq, Euronext, or the New York Stock Exchange. Crypto exchanges are slower in executing trades and broadcasting information about current prices and orders, which hinders quick dissipation of arbitrage opportunities. Consequently, arbitrage opportunities can be more frequent and last longer.[21] This suggests that crypto exchanges still need more time to achieve full maturity.

Indeed, the exchange landscape is still quite young and very dynamic. We see new entrants competing with longer established exchanges, often successfully, leading to frequent changes in the ranking of the most active exchanges.

In the first years of Bitcoin's existence, the most important crypto exchange was Mt. Gox, a Tokyo-based exchange. By some estimates, Mt. Gox was at one point responsible for handling 90% of Bitcoin trades. Mt. Gox's demise started in 2011, when the exchange was compromised by a hacker who managed to manipulate the site and the Bitcoin price

[21] See Gandal and Halaburda (2016).

it listed, and succeeded in sending him- or herself a large number of bitcoins, obtained at the artificially depressed price. Mt. Gox recovered from the attack, but its temporary weakness caused it to lose market share to competitors.

In spite of its problems, Mt. Gox remained the dominant Bitcoin exchange until mid-2013. In early 2013 it became difficult for U.S. customers to access Mt. Gox. Historically, U.S. customers were served using a bank account that belonged to a Mt. Gox subsidiary, but in May 2013 that account was shut down by the FBI. In February, 2014 Mt. Gox was again attacked by hackers, with an estimated $350 million worth of bitcoins stolen, leading to the shutdown of the exchange.

After the Mt. Gox shutdown, the Bitcoin market was in turmoil; predictably, the exchange rate of the cryptocurrency versus traditional currencies fell. Nonetheless, the market proved to be remarkably resilient, and new exchanges mushroomed to fill the vacuum left by the disappearance of Mt. Gox. By July 2021, coinmarketcap.com lists more than 300 exchanges, with Binance, Coinbase, and Huobi Global being the main players. Modern crypto exchanges have gained a higher level of maturity and adapted to the arrival of new players: professional (high frequency) traders, who are allegedly responsible for most trades of crypto like Bitcoin or ether. Their activities have gone a long way to decrease arbitrage opportunities and increase efficiency of crypto markets.

In fact, many crypto exchanges have gone a long way since the times of Mt. Gox, and are now voluntarily complying with regulators. The first reason for that is that exchanges need to gain or regain trust from their customers. Failures, fraud, and theft at exchanges did not stop after the demise of Mt Gox.[22] Complying with regulations increases transparency—and thus reduces the scope of price manipulation by the exchanges themselves—but also forces exchanges to adopt security measures to protect investors. Economists have been advocating for a long time that making markets safe is a key component of their success.[23] The second reason is that some exchanges simply wanted to remain accessible to their customers. In 2018 several banks in the US and in the UK started to ban credit card purchases on crypto exchanges. Those decisions were presented as safety measures to protect banks' customers against

[22] See https://selfkey.org/list-of-cryptocurrency-exchange-hacks/.

[23] See Roth (2015).

highly volatile assets and possible fraud. Of course, many people suspected that this move was also meant to hamper the development of the crypto ecosystem, a potential threat to banks' businesses.

Ironically, crypto exchanges, which emerged from an environment opposed to the financial system, are now part of the mainstream financial system. For instance, Huobi Global, one of the major crypto exchanges has been a public company listed in the Hong Kong Stock Exchange since 2017, and Coinbase, a US-based exchange went public on the Nasdaq in April 2021.

We have discussed in this chapter a myriad of cryptocurrencies and tokens, their use, and the related infrastructure. These coins are either directly or indirectly based on Bitcoin's ingenious way of solving the double-spending problem. It turns out, however, that the design proposed by Satoshi Nakamoto may also serve other purposes than managing digital currencies. And hence applications of blockchain technologies inspired by Bitcoin are wider than just a payment system. We explore those wider applications in the next chapters.

CHAPTER 6

Smart Contracts and Blockchain

Among all the topics about the developments or promises that came with the emergence of cryptocurrencies the concept of smart contracts takes a particular place. Smart contracts considerably enrich the set of services that blockchains can offer, moving them beyond the role of a cryptocurrency's ledger.

The term *smart contract* was coined by Nick Szabo to describe, in his own terms, "*a set of promises, specified in digital form, including protocols within which the parties perform on these promises*." In other words, a smart contract is a program automatically executed upon delivery of a specified digital input. For instance, the automatic payment of a rent is a smart contract: most banks today allow their customers to set up automatic payments where one has to choose the date, the amount, and the payee. Such smart contracts are executed automatically as soon as some conditions are satisfied (e.g., a date for the payment of a rent). The notion of smart contract thus predated Nick Szabo. His contribution was essentially providing a clear and concise definition, and popularizing the concept. Also, it is important to note that smart contracts and blockchain are two different things: a blockchain can operate without smart contracts and, perhaps more importantly, smart contracts do not need a blockchain.

Smart contracts are closely related to the digitization of information and relationships between parties. This is because the automation of the execution requires both digital encoding of the contract terms and digital

H. Halaburda et al., *Beyond Bitcoin*,
https://doi.org/10.1007/978-3-030-88931-9_6

input of information to call the contract. Since Bitcoin blockchain is a pure digital environment, it is quite natural that it would also have smart contract capabilities. Smart contracts on Bitcoin, however, are very limited. The terms of the contract are encoded in an optional field of a Bitcoin transaction, and the coding language used by the Bitcoin system, called *Bitcoin Script*, turned out to be complex, difficult to use, and limited in its functionality. This is a serious hurdle for a wide adoption. Indeed, if a language is hard to use, mistakes and bugs are more likely to show up. Also, it requires specific skills that not so many people may have. Another serious limitation of Bitcoin's implementation of smart contracts is that the type of instructions it offers is limited and does not go beyond simple terms like not allowing to spend a certain amount before a certain date, or requiring the approval of two or more parties to sign off on a transaction.

6.1 The Rise of Ethereum

6.1.1 History

By now, it is common to associate smart contracts with Ethereum. But as we have seen, the concept of smart contracts existed before Ethereum, and their implementation does not require Ethereum, let alone a blockchain. Nevertheless, smart contracts on a blockchain are intimately related to the history of Ethereum. Not only because it was the first blockchain to offer the possibility to handle a rich set of smart contracts but also because Ethereum's very history has a lot to do with the pitfalls and successes of smart contracts.

We saw in Chapters 4 and 5 that Bitcoin's limitations very quickly led to a frenetic activity in the cryptosphere. Back in 2011 or 2012, new cryptocurrencies were regularly emerging, each one proposing to correct or improve some of Bitcoin's features. Similarly, Bitcoin's limited ability to handle smart contracts seems to be the main motivator for creating Ethereum. In early 2013, a then 17-year old Russian-Canadian programmer, Vitalik Buterin (Vitaly Dmitriyevich Buterin), excited about the potential of smart contracts on a censorship-resistant, permissionless blockchain, proposed extending the Bitcoin Script to a *Turing complete language*. A Turing complete language is a programming language that allows to program anything that is possible to program in any machine language. On the one hand, the early programming languages were not

Turing complete, as they did not allow to program some operations programmable in other languages. Bitcoin Script, given its limitations, is not Turing complete either. On the other hand, most of the modern general-purpose programming languages, like C, Java, or Python, are Turing complete. Introducing a Turing complete language to Bitcoin would allow for more flexible and complex smart contracts.

As we had seen in Chapter 4, changes to the Bitcoin protocol are very difficult to introduce. There is no organization or body to steer the development of Bitcoin towards one specific direction (and to impose it): any proposed change must be accepted by a majority of miners, something difficult to achieve when, as is often the case, there is disagreement about which direction to take. Not surprisingly, Buterin's proposal was met with resistance.

In response, by late 2013, Vitalik Buterin released a white paper where he proposed a new blockchain altogether—Ethereum. That blockchain not only includes a Turing complete scripting language, but significantly differs in many other elements from Bitcoin, with the intention to optimally support smart contracts. While Ethereum utilizes its native cryptocurrency, *ether* (with the sticker ETH), in its transactions, Buterin's proposal was not centered around a cryptocurrency (like Satoshi Nakamoto's) but more around a system to build decentralized applications with smart contracts and offer the ability to manage other assets than a native cryptocurrency (e.g., financial assets, property deeds, cryptocurrencies from other blockchains, etc.).[1]

The name itself, Ethereum, is a telltale of Buterin's vision, as it is directly inspired by the word *ether* (or *aether*), the element that was believed to permeate the universe and allows light to travel. Ethereum is aimed at being a medium for applications and transactions.

While building an environment for better smart contracts was the main reason that led to the creation of Ethereum, Vitalik Buterin's experience with attempted improvement of Bitcoin also influenced the governance structure of Ethereum. Progress in computers and software is fast. Any system (hardware or software) needs to be able to adapt rapidly. Tomorrow we may have new needs and opportunities. There is

[1] The auxiliary role of ether is visible in the way it is listed on the cryptocurrency exchanges. Even though the exchanges buy and sell units of ether, they list it as "Ethereum."

no doubt that Bitcoin is ill-designed when it comes to keeping up with technological progress and changing needs.

Eventually Vitalik Buterin was joined by Anthony Di Iorio, Charles Hoskinson, Miahi Alisie, and Amir Chetrit as the founders of Ethereum. During early 2014 Joseph Lubin, Gavin Wood, and Jeffrey Wilcke also joined the founders. The early development of Ethereum, initially financed by Anthony Di Iorio, was made through a Swiss company, Ethereum Switzerland GmbH. The main part of the development was funded through a crowdsale between July 22, 2014 and September 2, 2014. The funding of Ethereum is interesting in itself. It is indeed considered by many (and rightly) as the first Initial Coin Offering (ICO), before even ICOs got their name. Backers were asked to send bitcoins to a specific Bitcoin address, in exchange receiving an Ethereum wallet with a password, which would allow them to access their ethers once the platform launched. For the first 14 days of the crowdsale the price was 1 bitcoin for 2,000 ether, and after that the price would drop linearly until a final rate of 1 bitcoin for 1,337 ether. In total, 31,529 bitcoins were raised (worth almost $18.5 million at that time), in exchange for 60,108,506.26 ether. Thus, unlike Bitcoin, some coins were pre-mined before. An additional amount of nearly 12 million ether was also pre-mined, to be split equally between the team of developers and the Ethereum Foundation, before Ethereum went live on July 30, 2015.

6.1.2 Ethereum Is "Different"

Most of the cryptocurrencies that were created shortly after Bitcoin adopted the same design as Bitcoin; the main differences were mostly with respect to some parameters like the total supply of coins that would eventually be mined, the exact version of the hashing algorithm used for proof-of-work, and other technicalities like the size of the blocks.

Ethereum kept the basic structure of Bitcoin's design, namely the use of hash pointers to link blocks as devised by Haber and Stornetta (i.e., each block contains the hash of the previous block) and the fact that the system would be maintained by miners who would earn block rewards (newly created coins) and transaction fees. Apart from this, Ethereum was built from scratch. This permitted its creators to reach a design as close as possible to their objective: a virtual network that could run decentralized applications and be a fertile environment for smart contracts. In

the Ethereum system, this network is called Ethereum Virtual Machine (EVM).

Just like Bitcoin, Ethereum participants interact with the blockchain through addresses. However, Bitcoin and Ethereum differ in what is recorded on the blockchain. In Bitcoin, a user or a miner has to have an address to receive and send bitcoins. Since Bitcoin's main purpose is transferring bitcoins, addresses hold unspent Bitcoin transactions outputs, i.e., bitcoins sent to an address in some previous transaction but not yet spent. Moreover, Bitcoin's blockchain only records transactions. If an address has received multiple transaction outputs, the net of bitcoins available to spend from this address (i.e., a balance) is not directly recorded on the blockchain. Instead, one needs to look through the transactions and tally it up separately. This is what is called an *implicit state* of the blockchain. In contrast, Ethereum keeps an *explicit state* on its blockchain. That means that for each address, the current balance of ethers is recorded. As transactions are sent to and from this address, the balance is updated.

Another, much more important difference is that addresses in Ethereum can be accounts or smart contracts. An account is an address that sends, receives, and holds ethers, but does not include any additional code. A smart contract is an address that may (but does not have to) send, receive, and hold ethers, but that crucially includes code that is run by the miners when the contract is called. A smart contract address also stores the state of the smart contract. For example, if the smart contract is a token contract (e.g., ERC-20), that is, a smart contract that creates and manages tokens, the state of the contract will include the current balances of the token held by different users of the token. What is important is that once the smart contract is created on Ethereum, that is, it is recorded on the Ethereum blockchain, its address and the code cannot ever be changed. The state and ether balance, however, change with every transaction.

The term *transaction* also has a broader meaning in Ethereum than it has in Bitcoin. Simply *sending ether* from one address to another—similar to sending bitcoins—is one type of transaction. Another type is *creating a smart contract*. And the third type of an Ethereum transaction is *calling*

a smart contract (i.e., executing functions defined by the smart contract's code).[2]

Executing a transaction requires some computations. They are trivial for a simple ether transfer but can be very intensive—and expensive—when calling a complex smart contract. This is especially true since, for instance, smart contracts can call other smart contracts. With that, what looks like a simple transaction may require running extensive code. In order to execute the smart contract, the miner runs the code on his/her computer. The fact that the smart contract is executed on the miner's computer is a key element in the incentives at play in the design of Ethereum. Since running the code of the smart contract takes some computing resources, the miner needs to be compensated for that; otherwise, the miners would only include simple transfer transactions in their blocks. In order to compensate the miners for running these computations, users who post their transactions need to pay a fee, which in Ethereum involves *gas*. The more resources one needs to run the program (i.e., the smart contract), the more the miner needs to be paid. On the Ethereum platform computing resources are counted in *gas*, a unit specially created by the team that developed Ethereum.

In Bitcoin, it is clear from the posted transaction how much space in a block that transaction is going to take. Transaction fees are voluntary, and transactions taking more space need to offer higher transaction fees (lest miners will prefer to include a larger number of transactions taking less space). In Ethereum, it is difficult to know how much computation a transaction will involve just by looking at it. Transactions fees are still voluntary, but the miners need help in figuring out which transaction is worth executing given the offered fee. This is the role of the Ethereum gas. To each instruction in the code of a smart contract there is a gas amount that is needed to execute the code. For example, an addition and a multiplication require 3 and 5 gas units to be run, respectively. But other instructions can use much more gas, like for instance *ecrecover*, which is used to verify digital signatures. That instruction uses 3000 gas units. A simple transaction (i.e., sending ethers to an address) consumes 21,000 gas units. How many gas units each instruction requires has been

[2] Smart contracts consist of several functions, some of them used for opposite actions (buying, selling, recording, erasing). Hence, we don't really call a whole smart contract. The phrase is a shorthand for calling a function in the smart contract.

established in the first version of Ethereum, and maintained in all the updates.[3]

While the amount of gas required to execute a particular transaction is set in the Ethereum system, the sender of the transaction decides how many ethers to pay for each unit of gas that will be used, i.e., decides on the *gasprice* to offer. The miners will select transactions that offer the higher gasprice rather than more overall fee (which is not known until the transaction is executed), similarly as in Bitcoin miners will prefer transactions paying more per byte than paying more overall. Thus, the price of gas depends on the level of activity on Ethereum's blockchain—that is, depending on the demand for gas.

In order to avoid miners being trapped in the execution of a never-ending smart contract the sender of the message has to specify the maximum number of gas units that will be paid to the miner (so-called *startgas*). When the miners run the code of the smart contract they will count, during the execution of the program, how much gas they have "consumed" so far (and the consumed gas units will be subtracted from the startgas, showing remaining gas that can be used in this transaction). If the execution needs fewer gas than what has been specified by the sender's *startgas* then the smart contract is entirely executed, and the remainder is returned. Otherwise, the program will stop when reaching the limit—the gas will be paid to the miner, but the smart contract function will not execute. It is then safer to post startgas higher than the gas needed, but setting too high gasprice unnecessarily locks the posted gas value (startgas x gasprice) until the smart contract is executed, which may be costly. It is therefore important for a user calling a smart contract to have an accurate estimate of how many gas units will be needed.

The gas ecosystem is the way Vitalik Buterin envisioned the stability of Ethereum. By requiring users to pay miners according to how much computing resources their contracts need acts as a safeguard against pointless, long-winding transactions. It also incentivizes the creators of the smart contract to strive for resource efficient code, lest it may repel the users from using the smart contract.

Just as in Bitcoin, miners in Ethereum prepare their blocks containing transactions (simple transfers, contract calls, and contract creation), and compete to include their block in the blockchain. The transaction fees are

[3] A detailed list of gas consumption depending on the operation can be found here: https://tinyurl.com/t3wke2ba.

collected by the node that succeeds in adding its block to the blockchain, but all the nodes competing to add their blocks to the blockchain need to run the code called by the transactions included in the block, as this will permit to check the new state of the blockchain that needs to be reported in the block. Moreover, nodes that build on top of the recently accepted block need to run the code as well, to verify that the new state reported by the new block is indeed the state resulting from the transactions accepted in the block. The only node that is compensated with transaction fees is the "winning" node.

Aside from the transaction fees, the miners are also rewarded with a block reward. From its beginning in 2015 until 2021, Ethereum blockchain achieved consensus solely via proof-of-work. At the same time, Ethereum's stated goal was to eventually move to a proof-of-stake consensus mechanism. Preliminary steps towards that transition were made early 2021, but at the time of this writing the move has not been done yet.

As we discussed with respect to Bitcoin, the proof-of-work consensus mechanism naturally produces accidental forks. The forks are resolved by the longest chain rule. In Bitcoin, the blocks that don't make it to the main blockchain are orphaned—they are not recorded on the blockchain and the miners who mined them do not get any reward. In Ethereum, delays between blocks are measured in seconds: initially, the average delay was set to 35 s, but it is currently 15 s (it is 10 min for Bitcoin). Such pace of new block creation is bound to create a large number of forks, and thus a large number of miners whose blocks end up on orphaned branches. This is problematic because a high probability of not being compensated after finding a "good" nonce in proof-of-work could deter miners from putting the resources into maintaining the blockchain. This is why in Ethereum blocks off the main blockchain are *uncled in* rather than orphaned. The miners on the main blockchain link those so-called *uncle* blocks to their block (i.e., include their hash in the block). In return the miner on the main blockchain path who linked an uncle block gets some reward and the miner who mined the uncle block gets partial block reward. He or she, however, does not get any transaction fees.

A number of blockchain characteristics that are fixed in Bitcoin are designed to be flexible in Ethereum. This includes the block capacity, interval between the blocks, the block reward, and the supply of ether. In Bitcoin, the block capacity is set to 1 MB, and attempts to increase it resulted in a hard fork creating a new cryptocurrency, Bitcoin Cash. In

Ethereum, the number of transactions that can be included in a block is limited by the amount of gas they will consume in total. This is a more adequate way as sometimes simple transactions (taking little space) may involve running complex and costly code (and thus take more time). The main difference, however, is that miners periodically vote on whether to increase or decrease the gas limit of a block. It allows to balance the demand and supply for transaction execution resources.

Similarly to the block capacity, the supply schedule of new bitcoins issued is set in the Bitcoin protocol, and changing it would require a consensus of the miners and a hard fork. The Bitcoin protocol fixes the block reward, starting with 50BTC, and halving it every 210,000 blocks. After the reward is cut below one satoshi, no more block reward is paid out. The total number of bitcoins created under this scheme is about 21 million. Ethereum does not have such a fixed supply schedule and a resulting hard cap on the total number of ethers issued. The amount of ether paid out as the block reward is adjusted with the updates of the Ethereum protocol. Initially every block paid 5 ethers on top of the transaction fees, and in 2021, it pays 2 ethers. But those changes are not scheduled in advance. And thus, there is no limit on the total number of ethers that can be issued in the system. Similarly, to the block reward, the targeted interval between the blocks is changing with the protocol updates.

In principle, anyone can propose a protocol update to Ethereum. In practice the updates that are actually adopted by the miners are coming from the Ethereum Foundation, the Swiss-based foundation that was set up at the beginning of Ethereum development, in 2014. It is not unusual for cryptocurrencies to have "foundations"—entities who aim to coordinate efforts around the development and maintenance of a blockchain-based platform, without any formal power over the miners or users. Examples include the Bitcoin Foundation or the Peercoin Foundation. These foundations differ widely in their effectiveness. The Bitcoin Foundation was dissolved in 2015, while the Ethereum Foundation, thanks to its reputation and influence, continuously plays an important role in the evolution of Ethereum. Vitalik Buterin's active involvement in the Foundation may be a key factor in maintaining this reputation and influence. In contrast, Bitcoin's creator, Satoshi Nakamoto, ceased to be visible about two years after creating Bitcoin; his last message was posted on the forum bitcointalk.org in December 2010.

When a proposed Ethereum update lowers the block reward, miners may be hesitant to adopt it. Conflicting incentives create a coordination problem: each miner may prefer to stay with the old protocol as it offers a higher reward; but if most other miners adopt the update, he or she will not get the reward at all, as their blocks won't be recognized by the blockchain anymore. When this happens, the Foundation's informal authority acts like some sort of coordination device: miners are not obliged to adopt the updated version proposed by the Foundation, but anticipating that most (if not all) the other miners will do so, it is in each miner's interest to also adopt the new version, thereby implicitly accepting the new terms set by the Foundation. The role of the Foundation as a "moral authority" became even more apparent during the so-called The DAO event in 2016.

With smart contracts allowing for automated execution of agreements between independent parties, came the promise of automating the whole organization. It requires complex, multi-contract structures, but these are possible to construct on Ethereum blockchain thanks to the Turing complete programming language. The first type of such complex structures was deployed on the Ethereum blockchain in 2015, and gained real traction in 2016; these are decentralized autonomous organizations (DAOs). DAOs are based on a premise that the whole organization or corporation may be automated with smart contracts, can be managed by all members of the organization, and will not need to rely on a decisive power of human managers, who could abuse it or extract value. While DAOs come with different names, like Dash, Steem, or MakerDAO, one of them, started in April 2016 was ominously called *The DAO*. Like many other DAOs, The DAO had no management structure outside of the layered smart contracts, no director, let alone a board of directors. The goal of The DAO was to decentralize venture capital. The principle of The DAO was relatively simple. Investors were participating by sending ethers to The DAO smart contract, like for a crowdfunding operation, and receiving in exchange a certain number of tokens (100 DAO tokens for each ether) that would permit them to vote on which project submitted to *The DAO* should be funded. The profits from the investments (if any) would then be redistributed to the funders in proportion to the amounts they invested. The principle was simple, but its realization was quite complex, with multiple connected smart contracts and so-called

children DAOs. While The DAO, like other applications of permissionless blockchains, is aiming to appeal to everyone, few people have the expertise to scrutinize the code governing such complex smart contracts.

The sale of The DAO's tokens started on April 30, 2016, and by May 21 the crowdfunding raised 12.7 million ethers (worth about $150 million at the time). On June 17, 2016, a hacker managed to steal 3.6 million ethers (worth about $50 million at that time) exploiting several vulnerabilities in the code of The DAO's smart contract. It is important to note that the attack did not violate the code of the smart contract. Instead, the hacker was able to steal the funds because the loopholes in the code allowed for operations that most likely were not intended to be allowed. Some of those vulnerabilities have been identified a month before the attack, yet the call for action to fix them was not answered.

The hack exploited a measure in The DAO contract that was initially meant to protect minority voters: backers who disagreed with the funding decision of The DAO could retrieve their funds and fund another project. This *retrieve & fund* feature was coded in The DAO smart contract through a "split function" allowing a user to retrieve his or her ethers from the DAO fund, and deposit them into a *child DAO* for some period of time. After that time, the user was free to fund another project. There was the loophole: the split function would first send the funds to the child DAO, and only afterwards update the user's balance with the main DAO. The attacker managed to repeatedly call this split function in a way that the function would never be executed entirely. Each time the function was called, funds were deposited to the child DAO, but the function was stopped before updating the balance with the main DAO. Thus, the attacker withdrew multiples of the funds he had with The DAO.

The theft showed in a painful way that a complex code, even if visible to everyone, may be too difficult to scrutinize; and even if issues are found, fixing them may be too slow or never happen in a fully decentralized, leaderless world. For a smart contract ecosystem like Ethereum, that was just taking off the ground, it was a serious blow and risk for its future. The theft also presented a dilemma: Some members of the Ethereum community, although annoyed by the attack, maintained that the blockchain should be ruled by the code, as the only objective reference, and considered that nothing should be done, taking it as a lesson when designing smart contracts. But many other people disagreed, calling for a return of the stolen funds to the original owners, on moral grounds. Eventually the latter group "won" the debate and the

unthinkable happened: Ethereum Foundation decided to re-appropriate the funds.

To reverse the theft, new transactions were added where the victims of the hack got their ethers back. The reimbursement, however, was not issued by the thief. Such transactions are technically not valid and under normal circumstances Ethereum mining software would flag and reject them. A software update issued by Ethereum Foundation in July 2016 allowed these transactions to be recorded onto the blockchain nonetheless.

The miners would accept the updates voluntarily. A significant number of miners maintained their opposition to the idea of re-appropriation of stolen funds. After all, it was contrary to the idea of immutability of the blockchain, a warrant of its security. Those miners did not accept the software update proposed by the Ethereum Foundation and simply continued to mine blocks on the historic blockchain, i.e., the blockchain in which The DAO hacker keeps his/her/their booty.

The majority of the miners, however, accepted the Ethereum Foundation update; and thus, on the main Ethereum blockchain the funds stolen in The DAO attack have been re-appropriated. The update was a so-called *hard fork update*, which means that the miners who did not accept the update cannot participate in the main Ethereum blockchain. Instead, the blockchain that is still building on the original Ethereum without this crucial update is called *Ethereum Classic*.[4] On Ethereum's blockchain the fork splitting Ethereum Classic from Ethereum is still visible, on blocks #1920000 to #1920009.

The existence of forked blockchains is not unique to Ethereum. As we have seen in Chapter 4, Bitcoin itself has experienced several such forking, most famously perhaps with the creation of *Bitcoin Cash* in August, 2017. All these forks have in common that they are due to disagreements between miners: some miners want to keep maintaining the blockchain with the same software and others want to update and add new functionalities that are not compatible with the old software. Blocks created by the new version may not be "readable" by the previous version but there is no issue with respect to the validity of the transactions in the two versions of

[4] Many of the updates in Ethereum are hard fork updates. But except for this one, they have been accepted by all miners, so they did not result in other, alternative versions of Ethereum.

the blockchain. Ethereum's hard fork is distinct in that one version, which many deem as the official version, does contain invalid transactions.

The fact that the overwhelming majority of miners followed the Ethereum Foundation's lead on this contentious issue speaks both to the moral authority of the Foundation and the importance of community agreement in the proper functioning of blockchains. The purported immutability of record history on blockchain is not a technical property, but the outcome of coordination of individuals who make decisions according to their own objectives.

There is no doubt that this fork, however, well-intentioned it was, is a serious dent to the notion that the blockchain assures an immutable record of history. Moreover, The DAO event was also a strong warning for smart contract enthusiasts and proponents: an ill-design contract may have dramatic consequences.

6.2 Smart Contracts

To understand what smart contracts can and cannot do it is important to understand in detail how they work (without entering too much into the nitty–gritty computing jargon). Reviewing what smart contracts look like on Ethereum will also help us to dive a bit deeper in how Ethereum works. Even though there are different blockchains, smart contracts work similarly on all of them.

6.2.1 Smart Contracts on Ethereum

As we explained at the beginning of this chapter, a smart contract is a computer code that, in the case of Ethereum and other similar platforms, is put "on the blockchain." Smart contracts are "put on a blockchain" by one of three types of Ethereum transaction—the *contract creation* transaction. After creation, a smart contract is an Ethereum account, with its address, state, and balance. Another type of transaction—*contract call* transaction—may be used to execute a function in the smart contract's code. Often, the function will take some parameters; they would need to be included in the contract call transaction. When a smart contract function is executed, the state of the smart contract is changing. At the same time, the code of a smart contract never changes. That makes it difficult,

for example, to fix a bug if it is found after the smart contract is put on the blockchain.[5]

Since the Ethereum programming language is Turing complete, smart contracts offer wide possibilities. Smart contracts can be used, for instance, to track shipments. As an illustration, consider a parcel sent from, say, London, England to Bernalillo, a suburb of Albuquerque, New Mexico. The parcel is likely to make the trip in several legs, say, first from London to New York, then from New York to Atlanta, then Atlanta to Albuquerque and then to Bernalillo. At each step the parcel would be scanned, and the result of that scan is sent to a smart contract, which in turn will send an update to someone (or simply update the database of the carrier). This looks very familiar because this is actually what carriers like Fedex or UPS are doing. The only novelty here is that a smart contract on Ethereum that tracks shipments could make it easier to track a shipment carried by different carriers. Tracking services by Fedex or UPS can only operate with the parcels they handle themselves.

Another example of a possible smart contract use is weather insurance. In such a contract, the payment is issued to a receiver *automatically* when, say, rainfall or temperature exceeds certain parameters (is too high or too long for a pre-specified length of time).

In fact, as long as one can write a program a smart contract can be about almost anything. While the term "contract" suggests several parties committing to undertake certain actions, this is not a requirement for a "smart contract." Some people have written smart contracts that are calculators or alarm clocks. To what extent such smart contracts are useful is a different debate; the point is to stress that, if it is programmable, then it can be a smart contract. Since the only limitation is whether a task is programmable, smart contracts can be also far more complex than our examples. For instance, a smart contract may include triggering the execution of a function in another smart contract. This is what we see in the development of many recent dapps (discussed in more detail in Sect. 6.6).

While creating a smart contract and calling a smart contract are transactions in Ethereum, similar to the transfer transaction (sending ethers), smart contracts are fundamentally different from moving cryptocurrency on blockchain. One of those fundamental differences is the need for a

[5] The way around it is to refer to external libraries that can be changed.

smart contract to connect to the world outside of the blockchain. In contrast, ethers and bitcoins are self-contained in their blockchains. Smart contracts would have very limited usability if they were similarly contained on the blockchain. Let us use our two previous examples to illustrate the issue.

Consider a smart contract on Ethereum that tracks shipments and see what happens when our parcel arrives at, say, the Atlanta facility. There, the parcel is scanned and some data from outside of the blockchain is sent to the smart contract (e.g., the reference number of the parcel, the word "Atlanta" and the date and time at which the parcel was scanned). Sending this data takes the form of a transaction, pretty much like a transaction between two wallets is made on Bitcoin. The only difference is that instead of sending bitcoins or ethers to an address we are sending here some data, and the recipient's address is in fact the smart contract's address.

That transaction called a function in the smart contract which, once executed with the inputs from the scanner, produced a new message. That message will be stored on Ethereum's blockchain and is simply the output of our transaction. But how do we get notified, that is, how do we get that email telling us that the parcel arrived in Atlanta? To do that, we need a computer that checks the blockchain regularly, searching whether the output of the smart contract shows up on the blockchain. As soon as this message is found, the computer sends us the long-awaited email. There is an important take-away here: smart contracts on Ethereum cannot execute actions outside the blockchain. Here, sending an email is an action that takes place outside the blockchain. The only way to do that is to have a system, external to Ethereum, that can read Ethereum's blockchain. Similarly, the smart contracts on Ethereum need external digital input, and need to be called externally. While Bitcoin blockchain can be fully isolated from the external world, and the only input it needs is transfer "disposition" signed with a cryptographic private key, a smart contract on blockchain in isolation would be severely limited in its functionality.

The most prominent vulnerability of smart contracts are potential coding mistakes (so-called *smart-contract risk*). The DAO event is perhaps the most famous example of what can happen when a contract is not well written. But the need for the external inputs and execution systems creates additional problems and vulnerabilities. Consider now our second example of weather insurance. The smart contract will pay out, say, when

the rainfall is above a certain threshold. The parties to the contract need to agree on the source of the information on the rainfall.[6] It could be, for example, a reading from a designated sensor or from The Weather Channel website. These information sources are called *oracles*. Information in these examples is signed with digital signatures of the sensor or the website API. The contract can be called by providing information signed by a source designated as acceptable oracle by the smart contract. If the information satisfies the conditions, the user then gets the insurance payout. In our tracking example, the oracle is simply the scanning device recording the location of the parcel.

The need for oracles creates so-called *oracle risk* for smart contracts. There are two problems with oracles. One is that they potentially can be manipulated. One of the earlier smart-contract based weather insurance schemes failed, because the users were pouring water over the local sensors, and collecting payout from too much rainfall. The second problem is that pulling this information "manually" from outside sources is often complicated, cumbersome and prone to technical failures.

New solutions are introduced to overcome these problems. For example, *oracle contracts* are smart contracts that focus on putting external information on the blockchain. One of the advantages of this system is that it gives to the data that has been put in the blockchain by an oracle the same level of immutability that monetary transactions have. Another advantage is that the data can be more easily and reliably used to call other smart contracts. If several smart contracts need the same data, that data does not need to be pushed onto the blockchain again. The data stored in an oracle contract can be shared, thereby making the system even more efficient.

At the same time, many designs emerge among oracle contracts that aim at decreasing manipulability of the oracles, mostly attempting to incentivize multiple sources of information and multiple validators of this information to participate. However, while with their continuously improving designs they alleviate the risks, they do not eliminate them.

These examples highlight an important feature of the smart contracts' ecosystem: one needs bridges between the blockchain and the real world. An important consequence is that the security offered by Ethereum does

[6] Typically, one party codes the accepted source of information into the insurance smart contract, and the other party agrees to the source by buying insurance from this smart contract.

in part depend on the security of the oracles. It may be difficult (if not impossible) to hack Ethereum, but it may be easier to compromise an oracle. If such a thing occurs then one may execute a contract that should not be executed, or not execute it when it should be executed.

The security of a smart contract is thus more difficult to ensure once it relies on external data. And this is a separate vulnerability from the smart contract risk mentioned earlier—the risk that a smart contract contains bugs or loopholes, like any program. Writing a contract that is bug free and that does not contain any loophole requires certain skills that not every person has. Connecting a smart contract to a reliable oracle requires additional skills and knowledge. So, smart contracts are not as "democratic" as it sounds in the sense that, while everyone is *allowed*, not everyone is *able* to compose a smart contract and submit it to be included in the blockchain. Not everyone is also able to assess whether a smart contract already existing on the blockchain is well coded and refers to reliable oracles—this is reserved only for those who have sufficiently good programming skills and knowledge, or those who can afford to hire people with these skills and knowledge. This also creates an opportunity for a rise of intermediaries that may provide standardized smart contracts, and stake their reputation on the correctness of the provided smart contract.[7]

Aside from the risks and vulnerabilities discussed above, the nature of smart contracts as a computer code results in other limitations, which are often overlooked or misunderstood. It would be a mistake to think that smart contracts make the very act of contracting or negotiating easier. Many contracts are difficult to agree upon and design, and usually require the skills of experts. Smart contracts are all about automatization of execution of the agreement. Such automation may have an impact on when the parties find such a contract attractive. But it is difficult to see how smart contracts would make negotiation and overcoming fundamental differences between incentives of the contracting parties easier. For example, it has been proposed that one could well have a smart contract with a negotiation protocol. For instance, a smart contract could be used between a seller and a buyer, where each party would send to the smart contracts a proposal (e.g., a price and/or a quantity). An algorithm inside the smart

[7] We are seeing a glimpse into this opportunity with the emergence of platforms offering a user-friendly way to set up and put on the Ethereum blockchain non-fungible tokens (which we will discuss in detail later in this chapter).

contract would help parties to converge to an agreement. But there is a caveat with such an approach: each party would need to agree in the first place about the bargaining protocol. Once an agreement is reached, it is not easy to map its terms into a reliable computer code—and that can make the process of contracting more difficult.

Finally, an important and significant limitation of smart contracts is that by their very nature their applicability is quite restricted. Since smart contracts require unambiguous digital input, such contracts can only be about "hard evidence," thereby limiting the scope of situations they can handle. They cannot handle ambiguity or intentions, which are common subjects of court considerations.

6.2.2 What Do Smart Contracts Need a Blockchain For?

The concept of smart contracts outlined by Nick Szabo predates Bitcoin by more than a decade. And since a smart contract is a computer code, blockchain is not necessary to run it. We can run smart contracts on a "non-blockchain" platform, as long as all the involved parties have access to it. Therefore, it is legitimate to ask what are the features or the advantages a blockchain provides in the context of smart contracts.

The pre-blockchain examples of functioning smart contracts include vending machines and automatic payment of credit card balance. Such smart contracts were constrained by what the operators, such as banks, proposed. In contrast, smart contracts on Ethereum allow any users to create a smart contract on any codifiable agreement. They are peer-to-peer and customizable. But such smart contracts could also be hosted by a centrally managed platform dedicated for hosting smart contracts. Whatsapp, Ebay, DocuSign allow for peer-to-peer interactions, with customizable content—messages in case of Whatsapp, goods for sale in case of Ebay, or documents for signing in case of DocuSign.

Blockchains also promise immutability: once uploaded onto a blockchain the smart contract would be there in unchanged form, forever. There are two aspects to this immutability: one is permanence ("forever"), and the other is invariability ("unchanged"). For the smart contracts to be on the blockchain forever, we would need the blockchain to be maintained forever.

Although blockchains and cryptocurrencies have gained a large support and popularity since the inception of Bitcoin, no one can say that such systems will still be around in, say, 20 or 30 years—which may be needed

if the smart contract encodes, for example, property deeds. To maintain a permissionless blockchain, we need a few enthusiasts to keep running it. As we have seen earlier in Chapter 4, for a permissionless blockchain to be secure and difficult to attack, we need a critical mass of validators to participate; and thus many permissionless blockchains "failed" by neglect. It seems that some blockchains like Bitcoin or Ethereum are more likely to resist the passage of time than others (and some have already disappeared like Karmacoin, discussed in Chapter 5), because they have a large user base and are among the most valuable coins.

Is a centrally managed platform that hosts smart contracts more likely to be around "forever"? Although the reasons why platforms cease to operate may be different than for permissionless blockchains, we must admit that it is difficult to guarantee that any platform has a better chance of maintaining operability in a distant future. History is replete with examples of platforms or standards that were dominant at some time but eventually waned. It suffices to look at the audiotape, or the VCR to understand that virtually no electronic technology, even if backed by solid and powerful companies, can pretend to be around forever. Similarly, once powerful platforms connecting peers, like MySpace, are not even recognized by name among the new generations of social media users.

The second aspect of immutability is invariability. One often understands that the records, including smart contracts, are not altered. Electronic records can always be altered, but well-functioning permissionless blockchain makes it very expensive to effectively alter records. In line with discussions in Chapters 4 and 5, by a *well-functioning permissionless blockchain* we mean a well-designed and popular blockchain, with a large number of validators and high value of the native cryptocurrency. Smaller value of the native cryptocurrency may open the door to a successful longest chain attack despite a good design, as we had seen with Ethereum Classic and Bitcoin Gold. Moreover, Ethereum's The DAO event showed that even in well-functioning permissionless blockchains smart contracts may be "canceled."[8]

[8] Recall that what a public blockchain solution can potentially offer is that any alteration will be noticed. Indeed, recall each block of data in a blockchain is interlocked with the previous block by containing the hash of that previous block. Modifying any data in a block will change its hash, thereby invalidating the interlocking of the blocks.

The invariability of a record as smart contract, or a call of smart contract is also secured by digital signatures. Recall that double spending in a permissionless blockchain, aside from being expensive and difficult, is only possible on transactions done by the miners (or in cooperation with them). We would expect a well-designed platform for customizable peer-to-peer smart contracts to make the signatures accessible. If the platform attempts to retrospectively change the code in the smart contract without the private keys of the creator, it will be easy to see that the smart contract is not properly signed. The users should not call such a contract. So even centralized platforms cannot change the smart contract imperceptibly and effectively. If altering of records is detected, centralized platforms and permissionless blockchains are similarly subject to reputation concerns and user loss.

The main difference between permissionless blockchains and centrally managed platforms (including permissioned blockchains) comes down to censorship resistance. In a permissionless blockchain restricting access is difficult to execute (it is possible if sufficiently many validators commit to black-listing of certain addresses). A centrally managed smart contract platform could restrict the rights of certain individuals to create or call smart contracts, or restrict the types of smart contracts allowed on the platform. Note that this is true whether the centrally managed platform runs a permissioned blockchain or another database structure. Hence the answer to the question of what smart contracts gain from a blockchain is censorship resistance. It's important to note that the benefit is not coming from blockchain, but from the permissionless nature of some blockchain systems. It's not about "blockchain vs not a blockchain," but about "permissionless system vs permissioned system." We discuss permissioned blockchains in more detail in Chapter 7.

6.3 Tokens

Tokens are another concept, aside from smart contracts that are popularized by Ethereum, and later other blockchains. While the term token has been used commonly for centuries, in the context of blockchain it went from a mundane term representing something auxiliary (as in "subway token") to an exciting term representing new possibilities (like utility tokens or NFTs). Many fungible tokens trade on exchanges in the same way as cryptocurrencies, but their technical design is different. Cryptocurrencies, like bitcoins, litecoins, or ethers, are transferred on their

own blockchains. Tokens, in contrast, are governed by smart contracts, and are supported by a blockchain hosting the smart contract (most commonly Ethereum). Because the smart contract can program functionalities and limitations of tokens, they offer a wider range of possibilities than cryptocurrencies.

6.3.1 What Are Tokens?

Tokens are created and managed by smart contracts, typically on Ethereum. Such a smart contract includes in its state a ledger that keeps track of which Ethereum address owns how many tokens, or which tokens. For example, the state in the smart contract

0xd26114cd6ee289accf82350c8d8487fedb8a0c07

that manages MyTokens would include information that the address

0xd2fc6738287b458797d8a9d4a1331f80a5daf73e

(we can call this address "Alice'') controls 3 tokens, and that the address.

0x11bc2f043bf8a63fcbc5dc6e4239635a13195cbb

("Bob") controls 1.7 tokens, etc. Tokens are transferred between users by calling a transfer function in the smart contract managing the token. After this function is called, the ledger in the state of the contract updates to reflect the new ownership. So after Alice calls function *transfer(Bob,1.5)* in the smart contract governing MyTokens, the smart contract's updated state records Alice controlling 1.5 tokens and Bob controlling 3.2 tokens. Now, calling the function *balanceOf(Alice)* will return "1.5," as this is the balance of the account.

Since the smart contract tracks the balances, the double-spending problem for such digital token is solved. It can be unambiguously determined how many tokens a given account controls (i.e., owns) at a particular time. An immediate consequence is that tokens can be traded pretty much like cryptocurrencies. The difference between cryptocurrencies and tokens is that tokens do not need to maintain their own blockchain. That removes one level of complexity—to make a new cryptocurrency functional there need to be users interested in transferring the

coin and also miners finding worthwhile to mine this coin rather than mining other coins (or doing something else altogether). A token "just" needs to attract the users, while utilizing the miners of the blockchain hosting the smart contract. That may also come with the risk—if the hosting blockchain fails, the smart contract governing the token will fall with it. This concern adds to the appeal of Ethereum—given its popularity it is unlikely that miners will stop mining on Ethereum.

Many tokens are issued based on a smart contract adhering to the *ERC-20* standard. Early examples of such tokens include Augur's REP, Uniswap's UNI, and Tether (USDT). The design of Ethereum is a collaborative project, and like with many such projects, individuals are making proposals, which are then commented on and debated. To manage such debate, the team behind Ethereum's development created a forum where collaborators could make and discuss proposals. When a proposal is first made, it is called an Ethereum Improvement Proposal (EIP). Once the proposal has been debated, finalized and accepted, it becomes part of the list of standards and is called Ethereum Request for Comment (ERC).

Ethereum was envisioned to be a blockchain hosting smart contracts and decentralized applications. Right from the beginning, tokens were considered a vital part facilitating smart contracts' and applications' functionalities. Thus, one of the first Ethereum Improvement Proposals made was to create a token standard. It was the twentieth proposal ever made, so it became EIP-20, and once accepted, it became ERC-20. This proposal was made by Vitalik Buterin along with Fabian Vogelsteller.[9] The standard consists of a list of functions that a smart contract needs to encode to be an ERC-20 token. The list includes—besides *transfer*, *balanceOf,* and a few other functions—a *totalSupply* function. This function specifies the maximum number of tokens that can ever be created in this smart contract. Thus, token-creating contracts that satisfy the ERC-20 standard have an explicit commitment on the total supply of tokens. Many users consider this commitment a necessary (but not sufficient!) requirement to guarantee that tokens can have some value in the absence of any reputation effects.

The ERC-20 standard also has a couple of optional functions, like *decimals*, which specifies how divisible the token is. For example, *decimals(4)* means that 0.0001 is the smallest possible fraction of the token.

[9] Vogelsteller and Buterin (2015).

Other optional functions are *name* and *symbol*, allowing for the name and symbol of the token to be written in the code. These functions improve usability of the ERC-20 token, but they are not necessary for the token to be considered ERC-20.

With a Turing complete programming language, there are many ways to set up a smart contract creating and managing tokens. They do not need to comply with the ERC-20 standard. Yet, today the immense majority of tokens are ERC-20 compliant. The dominance of the ERC-20 can thus be seen as some sort of decentralized standardization. It is a standard, but that has not been imposed by any centralized regulatory body.

What drives the adoption of the standard? There are a few factors. One is that the standard, serving as a template, makes it easier to set up a new token. But probably the most important factor is that adhering to the ERC-20 standard significantly improves the usability of the token on the exchanges and in the wallets. Because the functions used in an ERC-20 token are well specified and parallel to the same functions in other such tokens, the third-party exchanges and wallet codes can automatically and seamlessly adopt any number of such tokens. Without the standard, each token would need to be integrated separately to interact with the wallet or exchange.

At the visual level, the tokens associated with a particular Ethereum address are displayed in the same Ethereum wallet as ethers. So, Alice will see in her Ethereum wallet the 1.5 MyTokens along with other tokens she controls and her ethers. It is important that the address has some ethers besides the tokens. This is because selling the tokens involves calling a transfer function of the corresponding smart contract, and thus is a transaction on Ethereum, which requires paying gas in ethers.

At the same time, adhering to the standard does not mean that all ERC-20 tokens are the same, or that they are all safe. The standard wasn't established by any agency that can guarantee safety. It was proposed to facilitate interoperability with other elements of the ecosystem. While the standard specifies few functions related to transferring, the ERC-20 tokens may have extra functionality, with significant impact on the reliability of the token. For example, MANA, a token used for the virtual reality Decentraland platform, has additional "mint" and "burn" functions. If these functions are not programmed correctly, the funds will be appearing and disappearing unpredictably and the state of the accounts

will be misleading or outright incorrect.[10] Moreover, as many tokens may be only used in particular applications, the usefulness of those tokens depends on the usefulness of the application. Should the application fail or disappear the tokens will lose their utility and therefore value, pretty much like the miles one may have accumulated for an airline that goes out of business.

6.3.2 Use of Tokens

Programmability of ERC-20 tokens allows for different types of uses. It is useful to make the distinction between three categories of tokens, depending on their purpose: utility tokens, governance tokens, and security tokens.

Because tokens are programmable, they can be programmed to be the only way to access a particular service. For example, Filecoin is a blockchain-based peer-to-peer file storage system, which is accessible only with Filecoin tokens. We call these types of tokens "utility tokens." Like most other utility tokens, Filecoin tokens can be bought on exchanges, but they are also earned by the computers providing the space for file storage. The optimal design of utility tokens is a similar problem as for platform-based digital currencies, discussed in Chapter 3.

Another effect of programmability of ERC-20 tokens is that the tokens can be used for voting and other governance decisions. These tokens are called "governance tokens."[11] Entities external to the blockchain may set up governance tokens to facilitate voting. And tokens can also be used for governance decisions about protocols on the blockchain. The importance of smart contracts of on-blockchain governance is two-fold. Firstly, the governance tokens are set up via a smart contract. Secondly, in many cases, the changes proposed, vetted and then voted on through on-chain governance accessed by using governance tokens are applied automatically due to smart contracts. One of the oldest governance tokens is DAO, a token associated with The DAO, an autonomous decentralized organization discussed earlier that imploded in 2016 resulting in Ethereum Classic forking from Ethereum. Governance tokens are crucial for DAOs by their purpose and design. A more recent and successful example of a

[10] https://blog.chainalysis.com/reports/erc20-updates-march-2020.

[11] https://coinmarketcap.com/alexandria/glossary/governance-token.

DAO is MakerDAO, with its MKR governance token. MKR holders can vote to change the economic rules that govern the decentralized lending associated with the DAO, for example, on whether the protocol's debt ceiling should be raised. Existing governance tokens are typically traded on exchanges. New tokens may be sold—auctioned off to the highest bidder (like MKR), or they can be awarded to the most active users on the platform, as COMP, governance token of Compound Finance.

Another purpose for complying with the ERC-20 standard is to facilitate tradability. Even if a token is not promising access to services or governance, tradability opens up the possibility of using it as an investment, similarly to stocks and bonds. And thus most tokens traded on exchanges are "security tokens." The distinction between utility and security tokens is relatively vague, because utility tokens are also tradable, and their acquisition may be an investment, especially given that utility tokens may be sold (and traded) well before the development of the platform utilizing them. In fact, because of this possibility, utility tokens are often used for fundraising.

Issuing and selling tokens can be a quick way to finance projects upfront. The way to do so is similar to the one we can find on platforms like Kickstarter or Indiegogo. Users backing a project need to send some coins (like bitcoins or ethers) to a smart contract, and will receive in exchange a certain number of tokens. The money collected may enable them to work on the development of their project. As mentioned earlier, the very first crowdfunding where backers would receive tokens was Ethereum itself. Several months before the launch of Ethereum individuals could pledge some bitcoins and receive in exchange a certain amount of ethers (that they would be able to use only after Ethereum went live). Ethereum raised about $18 million this way. The success of Ethereum's crowdsale was followed by a large number of similar crowdsales, giving rise to a whole market for ICOs—initial coin offerings—which boomed in 2017 and 2018.

6.4 Initial Coin Offering

The majority of ICOs are set up on Ethereum. The basic structure of an ICO consists of putting a smart contract on a blockchain. To back a project, one needs to send some ethers to the smart contract, which sends back some newly created tokens. If the project involves creating a

new blockchain with its native cryptocurrency, the tokens bought in ICO are supposed to be later exchanged for the native coins.

There are several reasons why one wants to launch an ICO. First, one may need funding to run a team of software engineers and computer scientists. While the novelty of Bitcoin was enough to attract a few talented people who worked for free, most subsequent projects incurred substantial expenses, e.g., paying developers. Crowdfunding with tokens can be a quick and convenient way for entrepreneurs to raise some funds, especially in the world of cryptocurrencies where commitments and liabilities are hard to enforce, if not non-enforceable altogether. Obtaining funds from traditional channels like banks or investors can be more difficult and costly; the lack of history in this nascent industry makes it difficult for traditional investors to predict the potential success of a project. It is also more difficult for entrepreneurs to offer guarantees about potential success. As a result, they would offer fewer funds and/or require higher equity stakes, which could hinder the true potential of the projects. In contrast, going directly to the crypto-users would remove such constraints; an ICO, open to anyone and without any vetting by any financial institution, is clearly a faster and more promising way to raise funds than a loan or an investment in exchange of equity.

But crowdfunding has also its drawbacks. Given the novelty of the ecosystem, and without solid valuation frameworks, it may be difficult to distinguish between good and bad projects. It was therefore not surprising to observe a phenomenon that game theorists describe as pooling equilibrium: under uncertainty about the value of a project, proponents of less sound projects try to imitate the moves and appearance of potentially good projects. Indeed, very quickly proposals for ICOs had a similar look and feel: a slick website full of catchy sentences explaining that their project will revolutionize the world, together with a white paper outlining the technology behind the blockchain and the token (the consensus mechanism, the hash function, etc.), but very little regarding the value proposition. There is no doubt that many projects that managed to raise hefty amounts of money (in the millions of dollars) would not have passed a first inspection by business analysts. The failure or lack of success of several "large" ICOs show that a large base of backers does not necessarily mean a high value of the project. One reason why the "wisdom of the crowds" may fail in this case is that many ICOs got their backing based on hype rather than a careful analysis of the fundamentals. Such analysis often would not even be possible based on the limited amount of

information available about the projects. The confusion between popularity and soundness opened the way to a second, less noble, purpose of launching an ICO: the opportunity for a quick buck.

The third reason for running an ICO is that it can also give some information about the popularity of the project. Issuing tokens is then an opportunity to quickly establish a solid base of potential users. By observing how many users buy the token we can learn how many users are interested and likely to use our services. In other words, tokens are a little bit similar to a poll, but with a twist: the users buying the tokens are likely to eventually use the service.[12] After all, without any intent to use a particular service there is little interest in buying the corresponding token. Of course, this reasoning is valid as long as the purchase is not made for purely speculative purposes.

While many ICOs failed, few of them were a spectacular success in terms of the amount of money raised. A natural way to assess the success of an ICO would be to look at the Return on Investment (ROI), a standard tool in finance. It indicates how many percent the value of one's investment increased. Given that tokens and cryptocurrencies are highly volatile, ranking by ROI has little sense: it is likely that the ranking changes frequently. Instead, it has become standard to look at the amount raised by an ICO. Yet, even such a metric is not very informative about the success of the funded venture. As an illustration, let us consider the three most "successful" ICOs so far.

Under this metric (amount raised), the most successful ICO so far is Filecoin, a project aimed at decentralizing file storage (some sort of Dropbox on a blockchain). Nearly $257 million were raised for that ICO in September 2017. The general idea behind Filecoin is very simple. There is no doubt that in the future we will need ever more storage space for our files, and the steep increase of cloud storage is a perfect testimony of this trend. Filecoin is a project aimed at decentralizing the storage we need, by basically allowing anyone to share part of his/her hard drive to host files. Individuals would pay to get their files stored in the network (in Filecoin's token, FIL), and those payments would serve to reward those hosting files. This business model looks fine, but the nitty–gritty details make the problem more complex than it looks. Technical specifications like how the files are stored, the equipment requirements, the design

[12] See Bakos and Halaburda (2019).

of the consensus mechanisms, and other issues make the right balance between hosts, users and miners' incentives difficult to find. On the one hand, one wants that service to be as cheap as possible as it will allow for wide adoption. On the other hand, one wants potential revenues to be sufficiently high so that it attracts many people willing to host files on their computers and secure the participation of miners. Even though the ICO took place in 2017, so far the development is not finished yet. Some significant progress has been made regarding the (complex) design of the platform. However, in 2021, it is not sure yet whether the proposed Filecoin platform is economically viable.[13]

The second largest ICO is Tezos, which raised $232 million, in July 2017. Tezos is a general-purpose platform aimed at making the use of smart contracts easier. In particular, the design of Tezos is supposed to facilitate governance. But the history of Tezos was not a smooth ride. The development of Tezos was marked by disagreements between its founders and a class action lawsuit launched by US-based investors, accusing the Tezos Foundation and Dynamic Ledger Solutions (a company created to supervise the development of Tezos) of violating securities laws. In other words, the sale of Tezos tokens was allegedly an illegal offering of securities.[14] However, unlike Filecoin, the Tezos project went through, with Tezos' blockchain going live in June, 2018. Since then, Tezos has established itself among the main blockchain platforms, like Ethereum, providing services to well-known firms like McLaren Racing or Red Bull Racing.

The third largest ICO is Sirin Labs. This case is different from that of Filecoin, Tezos, or many other ICOs because the project of Sirin Labs is not to create a new blockchain or platform for specific applications. Sirin Labs is in fact a company manufacturing computers and cell phones that are especially designed to interact with blockchain applications. Those devices come with an Operating System that allows for users to store their wallets and access blockchain-based applications, a feature similar to an app store. The ICO was held in December, 2017, and raised about $157 million. The first "blockchain phone," called Finney, was released in December 2018. Although Sirin Labs is a different venture than Filecoin

[13] https://news.bitcoin.com/filecoin-miners-start-a-strike-fil-validators-claim-the-projects-economic-model-is-not-working/.

[14] Lewis-Kraus (2018).

or Tezos, its history was not a smooth ride either. Financial difficulties led the company to massive layoffs, and the company and its CEO faced several lawsuits. In spite of this, Sirin Labs managed to maintain its objective. However, the token launched by Sirin Labs has not been successful financially speaking. Introduced in January 2018, the token quickly spiked at $3, only to decrease to less than a penny.

6.5 Non-fungible Tokens

A different type of token that is becoming increasingly popular is the so-called non-fungible token (NFT). Such tokens differ from regular, ERC-20 type tokens in that each token is uniquely identified. There is one standard in Ethereum devoted to non-fungible tokens, ERC-721, introduced in 2017. A newer standard, ERC-1155 allows for both fungible and non-fungible (and semi-fungible) tokens. There are two main differences between fungible ERC-20 tokens and non-fungible ones, based, e.g., on ERC 271. The first difference is that non-fungible tokens are not divisible. While it's possible to own a fraction of an ERC-20 token, non-fungible tokens are owned in their entirety. At the technical level, the ERC-721 standard does not support the *decimals* function. The second difference is that the ledger for non-fungible tokens must indicate who is the "owner" of each token. It is the role of the function *ownerOf (tokenID)*, which is not supported by ERC-20, but is required by ERC-721. In contrast, the ledger for ERC-20 token indicates only how many tokens each account controls.

Similar to fungible tokens, non-fungible token standards were established with the purpose of facilitating the functionality of apps running on smart contracts—so called *dapps*. The ERC-20 standard is insufficient for tracking NFTs because each non-fungible asset is distinct and needs to be tracked separately. One of the most popular types of dapps are games, and this is where the ERC-721 standard was used the earliest. In fact, the very first game on Ethereum, CryptoKitties deployed in 2017, is centered around NFTs. In the game, one can trade and breed "kitties." Each CryptoKitty is an NFT, with its unique "genome." The genome translates to visual attributes, and some are more desired by the gamers than others—though only for aesthetic reasons, as there is no other functionality for the attributes. Two CryptoKitties can breed a new one, with its own new genome created from the genomes of the "parents." These virtual cats are bought and sold with ether. At the peak of the hype around the game,

CryptoKitties would sell for thousands of US dollars. The most expensive CryptoKitty changed owners in 2018 for 600ETH, worth at that time over 170,000 USD.

A more recent example of a game based on a similar principle is Zed Run, where the players can trade and breed digital horses (NFTs).[15] It has the additional functionality that the horses take part in races, and players can also bet on winners, like in regular horse races. Many more examples of games utilizing NFTs, with many different designs, abound. According to NonFungible.com, in March 2021 over 40% of NFTs were distributed in the gaming industry.[16] The second category was collectibles, like NBA Top Shots or CryptoPunks, a series of 10,000 unique 24 by 24 pixel pictures.

In 2020, NFTs became increasingly popular but the headline-making NFTs had a different raison d'être than CryptoKitties or Zed Run digital race horses. The main differences lay in the association of NFTs to off-blockchain items. As of today the most spectacular example is a token associated to a JPEG image, called "*Everydays - The First 5000 Days*" by Mike Winkelmann, a digital artist known as Beeple. The token was bought by Vignesh Sundaresan (a.k.a. Metakovan) for $69.3 million, in an auction run by Christie's in March, 2021.

CryptoKitties and Zed Run horses exist only on a blockchain, and while they can fetch hefty sums of money, they do not represent anything from outside of their respective games. However, as the tokens are programmable, they can include a hash of a digital file from outside of the blockchain. In this way, the token may be associated with this file. Such association is possible for both fungible and non-fungible tokens. Pre-Ethereum, there were experiments with Bitcoin colored coins to make the Bitcoins programmable, including association with outside assets. And in 2017 there was a series of collectible digital art files called Crypto Punks associated with ERC-20 tokens. However, the uniqueness and traceability of individual non-fungible tokens make them more suitable for such associations.

Since a token can be associated with an external file, many people would like to see such a token as representation of this file, allowing

[15] Note that Zed Run is on Polygone blockchain, which is related, but not exactly the same as Ethereum. See Lorenz (2021).

[16] Baloyan (2021).

for tokenization of digital art. Many people view the tokenization of digital art, or digital assets in general, as a tool to substantially simplify the management of such assets through smart contracts. For example, artists have been complaining for a long time that once a work is sold they completely lose ownership and any form of authority on it. Smart contracts are believed to bring some nuance or even solutions to that problem.

6.5.1 NFTs and Smart Contracts Do Not Solve Digital Art Ownership Problems

At the most direct level, an artist can sell an NFT associated with his or her digital art. But smart contracts can also allow for an ongoing "relationship" with the NFT even after it has been sold. A smart contract governing an NFT can be set up in such a way that the creator of the smart contract automatically receives a fraction of the price in all future sales. Or it can include a requirement that the creator of the NFT needs to "sign off" on the future sales, allowing the artist to exert a veto against any potential buyer. While such procedures can be set up with traditional contracting, tokens and smart contracts can automate them and thus ease the enforcement.

More generally, NFTs are considered by many people as an answer to a long-standing question regarding digital assets: is it possible to endow a digital asset with the same properties regarding ownership and possession as a physical asset? Those properties are easily defined and understood when it comes to physical assets for a simple reason: physical objects carry a notion of exclusivity. That is, if Alice has an object then Bob does not have it. If Alice gives this object to Bob then Alice no longer has it. In other words, any physical object is unique. When it comes to digital assets (i.e., information) this notion of exclusivity has no meaning. If Alice has a file, she can make a copy and give it to Bob while still being in possession of the file. This is what actually happens when we send a file to someone: our computers create a copy of the file and sends it to the recipient, without deleting our own copy. But there is more: the two files are exactly identical, there is no way to distinguish them. It is impossible to say that a file is the "original" and the other is a copy. This is why the management of ownership of digital information essentially boils down to copyright, patent, or trademark management. For instance, copyright is simply a way to distinguish between different copies: the person that has

the copyright over a song, an image or a text owns the right to decide the use that can be made of that song, image, or text. Other parties may be in possession of the work (i.e., the file), but without enjoying any right on it. Like for any physical object, copyright is exclusive and can be transferred. Non-fungible tokens are believed to be an answer to that problem. The idea is very simple: since an NFT enjoys, by design, the same exclusivity property than a physical asset, it would suffice to link a digital asset to an NFT to make this latter somehow exclusive.

Hitherto, the explanation we just offered regarding non-fungible tokens suggests that blockchain, smart contracts and tokens could, in the end, bring a significant improvement regarding the management of (digital) assets. This may be a hasty conclusion, though. The problem of digital asset management through a token is more complex.

The main source of this complexity is the following asymmetry: While NFT can be unambiguously linked to an external digital file via a hash pointer and the ownership of the NFT is also unambiguously determined by the ledger on the blockchain, the digital art may be linked to multiple NFTs, created by different entities not necessarily on behalf of the artist. In the absence of other forces controlling creation of NFTs, an NFT is a doubtful "representation" of the external asset. Such a problem is moot when the asset is in the blockchain, like a cryptocurrency or a token, because those coins and tokens do not represent anything outside the blockchain.

It is worth starting by clarifying what one gets when buying an NFT. If an NFT is linked to a piece of digital art, the art itself is not stored on the blockchain. What is often stored is metadata attached to the token that contains information regarding the asset. For tokens like Beeple's image the metadata includes Beeple's "signature," which presumably authenticates the author or the image. This signature is simply made using Beeple's private key and the hash of the token (which contains the hash of the image, too). The metadata also contains an Internet address where the image is stored. Since the information in the metadata is public, anyone has access to the original image created by Beeple (and can download it). In other words, the exclusivity property of an NFT does not transfer to the artwork file. Many digital assets that are linked to an NFT are stored on the *InterPlanetary File System* (IPFS), a peer-to-peer network where users can exchange files. The metadata associated with an NFT usually includes the IPFS address where the file can be downloaded. It is worth noting a major caveat: if the IPFS server hosting the file disappears,

the files it contained are no longer accessible. One can easily find on the Internet stories by people who bought an NFT and cannot retrieve the associated file.

Even if there is no problem with accessing the associated file, NFTs being sold do not include a transfer of copyright. That is, issuers of NFTs usually retain the copyright. At most, buyers get some partial rights to utilize the asset for their personal use, but not much more. But mostly, when they get an NFT, they just get a token. Moreover, given that Beeple retains the copyright, he can issue more NFTs associated with these assets.

While an NFT is associated with one particular asset, the same asset may be associated with multiple NFTs. Beeple could issue new NFTs associated to *Everydays—The First 5000 Days*. An even better example is another type of NFTs very popular in the beginning of 2021: NBA Top Shot. Those are simply short videos capturing some key moments represented by a token. The tokens are sold, pretty much like the token associated with Beeple's *Everydays—The First 5000 Days*. Like for Beeple's art work, those short videos are not owned by the buyers of the tokens. Those are just tokens. However, unlike Beeple, for each video the NBA is issuing several NFTs, differentiated only by a serial number. Those tokens do not bring much to the buyers except the fact that they own a token that can be resold (and each time this happens the NBA gets a cut of the transaction). In fact, if we think carefully those tokens are not much different from NBA or MLB cards. Owning a card does not bring anything else than the possession of the card. It gives no right or royalty over the image or the career of the athlete whose picture is on the card. That is, ownership of the NFT does not translate to the ownership of the associated asset.

Another hope around NFTs in the artworld is that they would help artists protect the authorship of their works, and thus serve as the certificate of authenticity. Since we can include virtually any type of information in the metadata that goes with an NFT an easy solution would be that creators add the digital signature of their work (e.g., by signing with their private key the hash of their creation) as soon as their work is created. In other words, an NFT could serve as a notary. This use of a blockchain would thus amount to going back to the original concept of the blockchain proposed by Haber and Stornetta: NFTs can serve to timestamp a document, and the author of a digital work would simply be the first one to create an NFT with that document. Two comments are in order. First, one does not need an NFT for that. It suffices to

attach to a transaction the hash of the digital work together with the digital signature of the author, something that even Bitcoin can support. Second, for this timestamping to work it has to be acknowledged by our societies (and especially our courts) that the first person to timestamp a digital work using a blockchain is the rightful author of the work. The problem is that a permissionless blockchain does not implement this rule. A (funny) example is a joke made by two economists, Mohammad Akbarpour from Stanford University and Shengwu Li from Harvard who, in April 2021, wrote a parody of an academic paper mocking the creation and the auctioning off of NFTs. In their (very) short paper they claimed that the paper would be linked to an NFT that would be auctioned off. To show that NFTs were not immune to predatory behavior and to question the value of legitimacy in NFT markets, Abdoulaye Ndiaye, another economist, from New York University, created an NFT of Akbarpour and Li's paper before they created their NFT.[17] As a result, there are now two different NFTs that are both linked to the same work (and since these NFTs are now in the blockchain they cannot be deleted). Apart from a good laugh among economists, the game played by these three economists leaves an unambiguous message: the creator of an NFT is not necessarily the creator of the digital asset it is linked to. And thus, while NFTs can infringe copyright, they cannot enforce copyright. Disputes related to authorship, authenticity or copyright will not be solved by an NFT, let alone a blockchain. We are thus back to square one: only third parties like courts can guarantee these properties.

These problems are driven by the *gateway problem* discussed earlier. The *Everydays—The First 5000 Days* token or the NBA Top Shot tokens are associated with assets that are *outside* the blockchain. The ownership of these tokens was never meant to represent the ownership of the associated asset. But suppose that we would want the ownership of the token to represent the ownership of the associated asset. For that we would need a guarantee that there is only one NFT associated with the asset (uniqueness), and enforcement of the ownership rights. Such connection between the on-blockchain token and off-blockchain asset needs to be created and protected off-blockchain. Thus, we could indeed have NFTs solving the

[17] https://mintable.app/collectibles/item/Economics-of-Non-fungible-Tokens-By-Mohammad-Akbarpour-and-Shengwu-Li/IS9Zm3L1ykjulKZ and https://mintable.app/art/item/Economics-of-Non-fungible-Tokens-Original-version-of-the-first-ever-academic-paper-on-NFTs/_HU2Kbal3JK-h6X.

ownership problem, but only if there is clarity on which creator's NFTs are honored and there is a guarantee that the creator is issuing a unique NFT for each asset. Moreover, the ownership of assets recorded by the NFTs needs to be honored by the parties involved, or enforced by the legal system.

To have art ownership through NFTs in the global art market, we probably would need some internationally recognized authority or set of authorities issuing these NFTs to secure uniqueness. But on the smaller scale NFTs can be useful without the need for such authority involvement.

For instance, NFTs could be used as tickets for events. The uniqueness of an NFT token makes it a perfect analog of a ticket with a seat number. Only the NFTs issued by the event organizer would be honored at the venue, and if the organizer issues multiple NFTs for the same seat, they all need to be honored (or issue a refund), just as in case of non-electronic over-issuance of tickets. What is more, the resale of such tickets can be made perfectly secure (the ticket is not delivered to the buyer until the payment has been made and the seller cannot cash in without delivering the token) and controlled by the token issuer (for instance, to prevent price gouging by ticket scalpers).

6.5.2 New Markets Enabled by NFTs

While high hopes are put on NFTs in terms of enforcing property rights for goods like digital art, our earlier discussion sheds light on important limitations of NFTs. Delivering on these hopes remains a challenge. At the same time, the event ticket example shows that there are other existing economic activities that NFTs can facilitate despite their limitations.

Interestingly, NFTs also create new markets. More notably, the new markets are markets for NFTs themselves, with all their limitations. We have already discussed examples of CryptoKitties and Zed Run—these are examples of markets for NFTs existing purely on blockchain. Many more such markets exist. F1 Delta Time allows users to buy, sell, collect, and race Formula 1 cars. These markets may be niche, and often are lasting for only a couple of years. But they can fetch quite large sums of money. Aside from the $100 K cryptokitty sold in 2017, some Zed Run race horses were bought for equivalent of $400 K in 2021.

Even though the NFTs like CryptoKitties are not associated with anything outside of their game, the technology behind NFTs allow for ownership of the NFT themselves beyond the game. Since the ownership

of an NFT is recorded on the smart contract on the blockchain, the owner can make use of it even if the game is no longer functional (display it, sell it, make use in another game—Ethereum language is Turing complete, so the possibilities are endless). In contrast, the day Linden Labs shuts down Second Life servers, players will lose access to all of their in-game goods.

With a notion of ownership independent of the dapp, NFTs thus blur the line between in-game objects and collectibles. This is why CryptoKitties, Zed Run, or F1 Delta Time emphasize "collecting" aside from trading, breeding, and racing. We have seen in Chapter 3 that markets for in-game digital items already existed. However, thanks to their non-fungibility and the impossibility to double spend, NFTs also allow for the first time for digital collectibles. NBA Top Shot is an example of pure digital collectibles, in the sense that there is no other function for them. So, NFTs created a new market for digital collectibles in the same way that baseball cards created a new market for… baseball cards. At the same time, unlike CryptoKitties, NBA Top Shots are associated with a digital asset outside of blockchain. (In a similar way that baseball cards are associated with players and teams.) The copyright and issuance of these NFTs is managed by the NBA.

There is also another new market that NFTs enable. Given that anyone could set up an NFT smart contract on Ethereum, there is interest in creating NFTs by individuals. There is also interest in buying such self-made NFTs. There are two obstacles for this new market to flourish.

Creating on Ethereum a well-functioning smart contract managing an NFT is tricky and requires specific knowledge. So, while the Ethereum technology allows everyone to set up such a contract, not everyone is able to set it up. This friction could prevent many people from creating NFTs associated with their files, holding back the supply. This is the first obstacle for the self-made NFTs to flourish.

The second one is achieving sufficient thickness of the market. Thick markets are essential for liquidity of any tradable goods, including fungible tokens. For non-fungible tokens achieving this thickness is more difficult, because the goods are idiosyncratic. Different items will appeal to different buyers' taste. So, it's important for the buyers to know where they can find items matching their taste. As there is more browsing than searching involved in the process of selecting an NFT for purchase, having them in one place is also important. Even if users would know how to create the NFTs on Ethereum, such NFTs would be difficult to find by the potential buyers in a completely decentralized world.

Fortunately, platforms and apps facilitating the creation of this market have already been created. Some even focus on NFTs associated with a particular type of outside asset. For example, Cent is a marketplace for NFTs associated with Twitter tweets.[18] Others are open to NFTs associated with an arbitrary digital file provided by the user. Examples here include OpenSea, Mintable, and Rarable, but many more are being created. They typically serve two functions, namely facilitating the creation of NFTs through an easy and user-friendly process, and providing a marketplace for those NFTs. Those marketplaces are, in many ways, similar to eBay, offering different sales options such as auctions or posted prices, facilitating the display, browsing, and searching. However, such platforms usually do not check for the copyright use of the associated files or the uniqueness of the NFTs associated with a particular file—just as Ethereum blockchain does not check for them. The lack of due diligence is what allowed Ndiaye to create an NFT associated with a file he did not have a copyright for (the PDF of the paper by Akbarpour and Li), and later allowed Akbarpour and Li to create an NFT associated with the same asset.

The need for market thickness in the self-made NFT market creates strong network effects (within a category). Buyers will go to the market offering the largest selection first, and will want to check out at most a few markets. Sellers will want to list their NFTs at the marketplace with most buyers. This two-sided network effect gives rise to a "winner-take-all" dynamic that typically results in a market with few powerful players. We have seen this dynamic in the first decade of the twenty-first century when Yahoo! Auctions and other competitors exited and eBay all but won the market for unique goods auctions.

Interestingly, NFT marketplaces face very similar problems as other online marketplaces, and they reach for similar solutions. Their business model relies on charging transaction fees. Some of them are charging for creating NFTs when the NFTs are set up, others are charging for it only when the created NFT is sold. All of them collect a fraction of the sales price. They rely on advertising and word-of-mouth to reach new users, and use strategic pricing in combination with network effects to attract sellers away from the competitors. Marketplaces with a larger number of buyers attract a larger number of buyers, and therefore can charge higher

[18] https://v.cent.co.

fees (as a percentage of the sale price) and still attract the sellers. Once one or two become dominant, the smaller ones can no longer charge fees high enough to be profitable and exit the market or fall into obscurity. So, while there are many marketplaces for NFTs created and more are being set up, the economic forces indicate that only few of them will stay in the market for the long run. The ones that do will be able to charge higher fees and extract the value created by the network effects.

Interestingly, a decentralized and open technology like Ethereum may give rise to platforms and intermediaries which are proprietary, centrally managed and capable of extracting value from the market. This is similar to the decentralized technology of the Internet giving rise to platforms and intermediaries like Amazon and Uber.

6.6 Dapps

Right from the beginning, Ethereum was envisioned as a platform for building flexible decentralized applications—dapps—with smart contracts. And thus, *dapps* are applications using smart contracts on a blockchain. Smart contracts are a necessary element of a dapp. And a smart contract accessible directly on Ethereum could be called a dapp. But it would have limited functionality and limited appeal to a wider audience. Typically, dapps involve multiple connected smart contracts (recall that a smart contract can call other smart contracts), and a user-friendly interface, like a website. Most often, dapps also have a layer (or a couple of layers) of additional software between the user interface and the underlying smart contracts.

The earliest dapps were games. The first game on Ethereum was CryptoKitties, created by Dapper Labs in 2017. It was also the first dapp to gain significant user traction. As mentioned earlier the purpose of the game is to buy, sell and breed digital cats, each with its unique 256-bit genome. The NFT smart contract complying with ERC-721 standard is one of the four interacting smart contracts which constitute the dapp. Other smart contracts govern the release of new Gen 0 cats, or auction mechanism for the cryptokitties put up for sale. Aside from the smart contracts, the CryptoKitties dapp has an off-blockchain algorithm governing creating unique genomes for the brand-new Gen 0 cats or for a new kitty, created by combining the genomes of its parents. Finally, the CryptoKitties dapp also includes the user interface allowing for easy buying, selling, breeding and displaying of one's digital cats. A user may

buy and sell his or her cryptokitties via the dapp interface. Then the sale is executed according to the mechanisms prescribed by the app, which includes a prespecified type of auction, and a transaction fee collected by Dapper Labs. But one can also sell the cats directly through the NFT smart contract. Dapper Labs still gets the prescribed fee (it is a part of the NFT smart contract), but the auction can be circumvented. It illustrates a general feature, since the smart contracts are on the blockchain. If a user can call the smart contract through the app, then he or she can call the same smart contract directly on the blockchain, if the smart contract address is known.

In our earlier discussion, we have brought up other examples of gaming dapps—Zed Run, F1 Delta Force; other examples abound. We have also talked about dapps facilitating digital collectibles via NFTs. CryptoKitties, aside from their role in the game, can also be considered a type of "collectibles" that are constrained to blockchain. Yet, most collectibles are associated with assets outside of the blockchain, like NBA Top Shots (created by Dapper Labs, the company behind CryptoKitties, in association with NBA), or digital art, like CryptoPunks.

But since the early days of 2017, there are now many more uses for dapps. Popular dapps include exchange marketplaces where users can trade their tokens, saving-and-loans contracts, payments mechanisms, and others. The self-made NFT marketplace which we mentioned earlier, OpenSea, is in fact a dapp. An even more diverse set of dapps is related to fungible tokens, the so-called DeFi dapps.

DeFi stands for decentralized finance. DeFi dapps provide a number of financial services, such as loans, insurance, crowdfunding, derivatives, and betting (prediction markets). They claim to "cut out the middleman" by relying on smart contracts. One example is lending and borrowing crypto. Dapps like UniSwap or Compound Finance offer borrowing and lending options via liquidity pools. Such pools use smart contracts to automatically set prices based on the supply and demand. As more token-owners offer a particular token for lending, the interest rate automatically decreases. When there is a shortage, the interest rate increases, attracting more token owners to offer their tokens.

On the borrowers' side, the interest rate also responds to supply and demand. But it does not need to be the same interest rate that the lender gets, as the liquidity pool may take the cut. The cut may be kept in the liquidity pool, to provide independent liquidity, or it can be cashed by the creators of the liquidity pool. A special challenge with

lending in a permissionless environment is that borrowers are anonymous. That makes it difficult to follow up with them and collect the amount due. Therefore, borrowers need to post collateral—typically in different coins than they are borrowing. And since the prices of coins are highly volatile, the pools require collateral that is multiples of the borrowed amount. It is not unusual to require collateral equal to 150% or 200% of borrowed value. Thus, the loans are not the "small business" loans that are offered by banks. Yet, they may be useful to facilitate short selling and arbitraging—-which may stabilize coin prices across exchanges.

One of the main arguments for the use of DeFi dapps is that dispensing with intermediaries or middlemen entails lower fees. The picture is more nuanced, though, because since a dapp relies on a smart contract on a blockchain it is impossible to use a dapp without paying any fee. Absent any substantial increase of the blockchain throughput, an increased dapp activity would increase congestion, which would translate in users paying higher transaction fees. Moreover, even aside blockchain transaction fees, dapps could be set up as a middleman and capture high fees themselves.

The term dapp stands for decentralized application. But the name can be misleading. While dapps are powered by smart contracts running on blockchain, even a fully decentralized and permissionless blockchain is no guarantee that the dapps running on it are also decentralized. In fact, while it is possible that a dapp is solely governed by smart contracts, like Uniswap, many dapps are centrally managed by their creators, like NBA Top Shot or CryptoKitties. This is because smart contracts themselves can be permissioned, with the contract creator exerting significant power over the participants in terms of pricing and potential changes. That may sound surprising given that the code of the smart contract cannot be changed after deployment on the blockchain. But smart contract code can refer to external libraries that can be controlled by the contract creator. Moreover, aside from the smart contracts, dapps also have layers of other software that can restrict access and increase prices. Thus, in environments with network effects once dapps gain sufficient presence, they can leverage the network effects to take advantage of their market power in the same way as traditional Internet platforms. Smart contracts and blockchain are not a panacea against powerful intermediaries.

One argument in favor of dapps successfully disintermediating traditional platforms is that anyone can set up a dapp, and that the code of the smart contract is visible on the blockchain (if the blockchain is transparent). A caveat in this argument is that the same logic is not sufficient

to defend against the rise of powerful intermediaries on the Internet. The barrier to replacing Uber or Facebook is not the difficulty to create a software with the same functionality or that the functionality is not known. The difficulty is to grow the network large enough that the alternatives are attractive. These intermediaries are protected by powerful network effects. The same could happen with dapps.

At the same time, decentralized dapps, governed solely by smart contracts, are much more effective than any decentralized applications and platforms before blockchains (like Wikipedia or open source software). Thus, it is more likely that the blockchain and smart contracts will complement the tools we currently have, rather than replacing them altogether.

6.7 Blockchain Governance, Dapp Governance, and Smart Contracts

A decentralized, peer-to-peer electronic currency requires an autonomous platform, that is, a platform that can somehow manage itself, without the need of a third party. This is particularly problematic when the platform consists of computer code and hardware because technological progress in computer science and electronics makes the update of any system inevitable. Also, the more complex a computer code is, the more likely that it contains some bugs or loopholes. The programs and operating systems we use every day are constantly updated not only to add new functionalities (which wouldn't be possible with older generations of hardware) but also to improve their efficiency and correct errors, bugs, or loopholes. The computer codes of platforms like Bitcoin or Ethereum are certainly not the most complex ones, but that does not protect them from the fate of virtually any computer code: any version will inevitably contain bugs and/or loopholes that need to be corrected and enhancements to adapt to new hardware possibilities. For permissionless blockchain platforms this is a crucial issue. Sooner or later the code will need to be tweaked, rewritten, corrected. And here comes the big question: how to upgrade such systems? In other words, what is the governance structure?

In this respect, the case of Bitcoin is instructive. Although some individuals have some sort of influence (i.e., their opinions are more heard than that of others), Bitcoin completely lacks any governance structure. Nakamoto was well aware of this problem, and considered that Bitcoin is designed as a democratic system where "voting rights" are simply defined

by the computing power of the miners with the now famous expression: *one CPU, one vote*. This voting procedure comes directly from Nakamoto's consensus mechanism, which aims at selecting which version of the blockchain will survive when, following a fork (accidental or voluntary), miners are confronted with different versions. If more than 50% of the miners work on the same version this latter will grow faster, entailing all the other miners who apply the longest chain rule to also mine on that version. If the system needs any update or correction, such changes could be debated and then submitted to a vote. The voting process is very simple. Miners only have to agree when they update the program running on their servers. If fewer than 50% do run such a change, that is, if the majority of voters/miners do not adopt this change, then it could be likely that this change will not survive.

There are several caveats in this reasoning. First, the one man, one vote principle is not equivalent to one CPU, one vote. This would hold if all miners had the same computing power (the same number of CPUs dedicated to mining if all those CPUs had the same computing power). In practice, this is far from being the case. The problem is even more severe because now the vast majority of the mining power on Bitcoin is held by pools that each gather thousands of miners (or servers). Second, this is a rather naive approach to voting. The formal analysis of voting procedures started in the second half of the eighteenth century with the works of the Marquis de Condorcet, a French philosopher and mathematician. Voting theory, which is part of the discipline known as Social Choice, was profoundly changed with a series of results in the 1950s and 1970s. A first, now famous result is the so-called Impossibility Theorem by Kenneth Arrow in 1950, which states that it is impossible to aggregate individuals' preferences over alternatives as long as one requires this aggregation to satisfy certain basic properties. Hopes for an ideal voting system took a hit with another impossibility result proved by Allan Gibbard in 1973, and Mark Satterthwaite in 1975. This result states that as soon as there are three or more alternatives, any voting system is either dictatorial (decisions depend on one voter only, the votes of the other voters have no impact whatsoever) or prone to strategic voting (i.e., absence of sincere voting). It is therefore no surprise to see that large agreements among Bitcoin miners about substantial changes in the Bitcoin software have been rare, if not non-existent. This is a pity because there is no question that Bitcoin's system is rather primitive: it does not allow for complex contracts like Ethereum, Tezos, or

other more recent blockchain designs, and its throughput is extremely low (preventing Bitcoin from becoming the world's currency), it consumes too much energy, etc. Most of the attempts to implement a significant update have failed, often due to disagreements about the way to update the Bitcoin software. Such disagreements fall right into Arrow's impossibility theorem: although there is apparently unanimous agreement about the fact that Bitcoin needs to be updated (not complete, some think that we should stick with Nakamoto's original design), the heterogeneity of preferences regarding how to update is such that it is extremely difficult, if not impossible to obtain a consensus about the update.

Bitcoin's lack of governance became quickly apparent, so that most (if not all) new blockchain designers after Nakamoto considered the question of governance. This question does not have an obvious answer for permissionless blockchains, for the simple reason that there is no way to enforce any change. Miners are free not to follow any recommendation and thus not to implement the proposed change. Also, the way updates are done does require the consent of the majority of miners. Indeed, platforms like Ethereum and others proceed to update by simply forking the blockchain. That is, at an agreed upon time miners are supposed to start mining with the new version of the software. Since there are always some miners who may disagree (or simply haven't updated their systems in time), the blockchain will fork with two branches: one branch continuing to mine with the old system and one branch mining with the new version. Updates manifesting as forks is not an ideal solution, as it fosters division and confusion in the community.

It has not been a problem for Ethereum, because of the Ethereum Foundation's "moral authority" in guiding the development of the system. When the Foundation announces the update, miners adopt the new version because they expect other miners to adopt it as well. This method works fine as long as the organization has enough influence over the miners—which makes the system somewhat centralized. But not fully so: if the proposed change is too controversial or if the organization would lose its "authority," smooth evolution of the platform may be more difficult to achieve.

So, is there an "impossibility" result regarding permissionless blockchain and ability to evolve? Perhaps not. The founders of Tezos have found an original and apparently promising way to proceed. The very structure of the Tezos software allows for updates to the protocol without the need to fork.

Updates to the protocol may mean the change in the design itself of the blockchain. Such design features include, for instance, the type of the hash functions, the structure of fees, the size of the blocks, the way "successful" miners are selected (proof-of-work, proof-of-stake, etc.). All those specifications can be understood as parameters, and could thus be encoded as variables in the blockchain itself. This is in contrast to Bitcoin protocol where all the design characteristics are hard coded. Any block that has been added in would then just need to be compliant with the latest specifications. But then, how are those changes decided? The answer found by Tezos is simply to put proposals for updates for a vote. Participants would cast their votes through a smart contract, and if the proposal receives enough votes, the smart contract updates the new "specifications." That way the blockchain design can be updated without the need to fork. This, however, does not eliminate the risk of a fork. Miners or participants who disagree with the changes can still fork the blockchain and continue mining with the old specifications. But at least Tezos has found a solution to the governance problem that is inherent to permissionless blockchains. Whether this solution will indeed be effective in the long term is something we will learn in the future.

While Tezos uses smart contracts for governance of the whole blockchain, many dapps use governance tokens and smart contracts for the governance of the dapp system. Few dapps are set up permanently without the possibility of future adjustments. The adjustments when they are allowed could be done unilaterally by dapp creators if the dapp is centrally managed, like CryptoKitties or NBA Top Shot. For dapps changes are proposed and voted on using governance tokens. Governance tokens may be bought (as MakerDAO's MKR), or may need to be earned (as Compound Finance's COMP). A user needs to have a minimum number of governance tokens to make a proposal for a protocol change. And then, the voting to accept the change is via smart contract with governance tokens. A user who has a large proportion of governance tokens may propose and vote in their proposal. This, again shows that dapps that are decentralized by design, may become centralized in practice.

CHAPTER 7

Enterprise Blockchains

The development of cryptocurrencies, together with the potential uses of smart contracts and tokens, led to expectations that the "blockchain technology" could be used for almost any situation. The type of blockchain technology goes from tracking diamonds to sharing medical data or management of royalties for artists.[1] Interestingly, many of the envisioned uses are letting aside the idea of tokens or coins by having a more general view about this technology. Blockchain is seen as a more general system of managing data, especially data related to value transfers, where double spending is an important vulnerability. A lot of this general view on blockchain relates to distributed databases, a well-known concept in computer science and in data management alike.

7.1 Distributed Database: What Is It and What For?

We have already seen that many aspects behind the Bitcoin design have been in fact around for some time, well before Nakamoto came up with Bitcoin's introduction. Hash functions, the notion of proof-of-work, and public-key encryption were tools or designs that were well understood

[1] https://everledger.io/industry-solutions/diamonds/, https://www.burstiq.com, and http://www.mediachain.io.

H. Halaburda et al., *Beyond Bitcoin*,
https://doi.org/10.1007/978-3-030-88931-9_7

in 2008 when Bitcoin was created. For instance, as we have already mentioned, the tamper-evident properties of blockchain—data linked by hash pointers—was already explored in the work of Haber and Stornetta in the early 1990s. Though the term "blockchain" was coined only after Nakamoto announced Bitcoin.

Similarly, distributed databases have been around for a long time before Bitcoin. A primary reason for using a distributed database is to avoid technical failures. If there are multiple servers hosting the same database (i.e., the database is distributed), the risk of the data not being accessible is reduced compared to the case when only one server is hosting the database: if one of the replicated servers fails, the data is accessible on the others. Data becomes inaccessible only if all servers fail simultaneously. Scale is another reason why one might be interested in distributed databases. If the number of requests made by users is extremely large, distributing these requests across several servers may mitigate congestion and thus allowing for a timelier service.

In fact, we all interact with distributed databases on a daily basis, when accessing search engines like Google or large social media sites like Facebook. The amount of traffic for such websites is so large that several, nearly identical servers are needed. No server in the world would be able to handle all search requests that these services receive at every moment. The solution is then to have multiple servers replicating the same data, and each time a search query is made, a request is dispatched to one of the hundreds of thousands—or more-—of servers Google has.[2] It would not be desirable if the results of a search query would be significantly different when handled by different servers. Thus, Google must make sure that all its servers have the same information, i.e., the same database. Since the information available on the Internet is constantly evolving, Google needs to continuously update those databases. Updating multiple servers in a consistent way presents a challenge. Transferring data between servers takes time, and if the data updates arrive often and with unpredictable delays, then well-established results in computer science show that we cannot guarantee that, at any moment, all servers have exactly the same copy of the data.[3]

[2] The number of servers is not public but estimated to be around 2.5 million! See https://en.wikipedia.org/wiki/Google_data_centers.

[3] This impossibility has been established by two famous theorems in computer science, the so-called FLP and CAP theorems. The FLP theorem was proved by Fischer, Lynch

The problem of updating data is not merely a technical problem. To see this, suppose we upload a photo on Facebook. Our computer (or phone) is connected to one of Facebook's servers. That server will be the first one having the photo and will start, almost immediately, to broadcast this photo to the other servers. If some of our friends consult our Facebook feeds shortly after we uploaded the photo, some may see the photo and others not. This is because some of our friends would be connected to a server that already has a copy of the photo and other friends are connected to a server that hasn't received the photo yet. Until now we could think that this is not an important issue because, after all, we are only talking about a photo on Facebook.

A slightly more complex problem is the following. People are using more and more online suites like Google Docs. Such services may not have all the functionality that a standalone program may have (and with which we can work offline), but they allow users to work together, at the same moment, on the very same document. However, what if, at roughly the same moment, two collaborators do contradictory manipulations on the document? That is, what if one collaborator is changing the typeface of a sentence to, say, italics, while the other collaborator is changing some words in the sentence? Google's servers will receive conflicting information. One server will put the sentence in italics and the other will change some words. What is the final outcome? To solve such situations Google needs (and has) a protocol to resolve conflicts, pretty much like with Bitcoin the longest chain rule is the protocol that miners follow when they face conflicting versions of the blockchain. In the case of simple services like Google Docs the version that will remain is simply the most recent one, that is, the latest that Google has received on its servers. The loss of information is not an issue because users always have the ability to use the version control options and see the history of changes. Similar conflicts arise for file storage services like Dropbox. If Dropbox's servers receive two conflicting updates (i.e., the same file modified on two different devices), then Dropbox will keep the most recent one and make a copy of the other file, marking it as "conflicting copy." So, no information is lost.

Consider now an even more complex problem: banks or credit card companies. Many banks have a large number of users and thus might need

and Patterson (1985), and the CAP theorem was conjectured by Brewer (Fox and Brewer, 1999), and proved by Gilbert, Lynch (2002).

to distribute their traffic and operations on several servers. The problem is that such institutions cannot afford to have conflicting entries like Google Docs because there can be only one version of our bank account or credit card balance. If, for instance, someone has only $100 left on her bank account, but makes at the same time two purchases of, say, $80 each, then only one of those operations can be recorded. One may argue that this is why some banks allow some of their customers to have a negative balance. But that only defers the problem. What if someone has an authorized negative balance of up to −$1,000 and at the time of making two purchases of $80 has a balance of −$900? The problem of conflicting entries is not solved. Banks and credit card companies have of course invested heavily in fast and very reliable network architecture to minimize the number of times such conflicts may occur. But the problem is not fully eliminated. The impossibility to have, at any time, perfectly synced databases is a fundamental property of distributed systems that one cannot get away with.

But isn't Bitcoin's double-spending problem similar to the banks' overdraft problem? Bitcoin has a way to resolve conflicts that we have already commented on in Chapter 4. If a user sends two conflicting transactions, then only one of them will be selected by miners when constituting a block. If two miners find valid nonces to their blocks at the same time and have selected different transactions, then the blockchain will fork. Due to the longest chain rule, however, eventually only one version of the blockchain will survive, and thus only one of the conflicting transactions will be recorded. Yet, despite the success of Bitcoin in achieving a reliable decentralized ledger of value transfer, it would be difficult for banks and other established commercial enterprises to benefit from implementing such a blockchain.

7.2 Limitation of Bitcoin's Blockchain for Enterprise Use

Given Bitcoin's blockchain success in managing cryptocurrency data consistently among a large number of independent parties, it is natural to ask whether a similar blockchain design would be useful for businesses in other contexts. The need to process big data in multiple servers and share the data between entities is present for many companies, not just the giants like Google and Facebook. So, on the

surface, the Bitcoin blockchain seems to be perfectly addressing businesses' needs. A closer examination, however, shows that the capabilities of the Bitcoin blockchain come with some serious caveats. To understand the caveats, it is important to remember that Bitcoin's blockchain is tailored to handle "monetary" transactions. As we have already seen in Chapters 4 and 6, the immutability and security properties that Bitcoin or Ethereum blockchains enjoy come from an economic incentive scheme ensured by the high value of their native cryptocurrencies. Those incentives rely on costly mining and valuable rewards through the mining reward. It is not clear how blockchain without a valuable, native cryptocurrency can induce the same incentives.

There are several aspects of the Bitcoin blockchain design that can make this technology less attractive for enterprise use. One is Bitcoin blockchain's low throughput and long transaction delays described earlier: the lack of trust among parties implies that transaction (or data update) validation can be slow. Not only blocks are added every 10 minutes, but also it is recommended to wait six confirmations in case accidental forks occur. Customers would not accept to wait one hour for a credit card payment to be accepted by a seller. The alternative to waiting would be to accept probabilistic settlement (or data update). That is, taking the risk that another longer branch of blockchain may show up and invalidate the quickly executed transaction. To make things worse, Bitcoin's or Ethereum's settlement is only probabilistic even after waiting for the recommended number of confirmations; it is only that the probability of another branch replacing the blockchain drops significantly. Clearly, no bank or credit card company would find such probabilistic settlement or long wait attractive.

Another issue is the high cost of operating the system. Bitcoin and Ethereum (at least for now) rely on the costly proof-of-work to achieve security. They are "self-sustainable" because hefty rewards obtained by the miners pay for the investments and the energy needed for mining. As Bitcoin's price exceeds $50,000, the 6.25 bitcoins block reward awarded to miners every 10 minutes is worth over $300,000. It adds up to $45 million a day. If an enterprise was to set up a similar blockchain system, it would need to pay the miners a similar amount to achieve the same level of security. Few enterprises would find it attractive.

The high cost of the system, as well as the problems with delays, scalability, and probabilistic settlement, arise because Bitcoin's system is

set up to achieve consensus in a permissionless environment, where no entity trusts another. It turns out that while blockchain has many features appealing to businesses, the complete permissionless and lack of trust is not as important.

7.3 Consensus for Permissioned Distributed Systems

The fact that Bitcoin's design is not adapted for enterprise use does not mean a blockchain has no use for businesses and organizations: a blockchain does not need to be designed in the same way as Bitcoin. Since Bitcoin is permissionless (i.e., anyone can have access to its blockchain and become a miner), its design is based on the implicit principle that participants cannot trust each other. For permissionless blockchains, the absence of a central authority that can control the platform—for instance, to prevent malicious behavior—requires a design that can ensure the integrity and reliability of the system when participants do not know, and thus cannot trust, each other.

In contrast, the environment that most firms or organizations are facing is such that participants are usually known, and there is some level of trust between them, or trust in the enforcement of contracts by the law. If an enterprise can have some sort of control over which party it will interact with, and especially who has the writing access to the data, the lack of trust would not be as much a hurdle as it is with a permissionless system like Bitcoin. When there is the possibility to restrict access, the blockchain is called *permissioned*.

Permissioned blockchain can easily allow for a fine tuning of access rights. For instance, some data can be made accessible to only a subset of participants. Similarly, some participants may be restricted with respect to their ability to add data to the blockchain and the type of data for which they have this right. All those restrictions can easily be implemented using cryptography, pretty much like one cannot make a transaction from a wallet in Bitcoin without having the private key associated with that wallet. Permissioned blockchain thus make it easier to keep at bay participants that are likely to create havoc or tamper the data. Participants that have writing rights for permissioned blockchains are called validators.

Any distributed system requires protocols to ensure consensus, even in a permissioned setting. In fact, this is a well-researched problem in computer science. In distributed systems, a node that has new data needs

to send it to all the other nodes, and a major issue is to ensure that data updates across nodes are correctly relayed and received by all participants. Problems may arise when some nodes can crash and, thus, are unable to relay and/or receive data updates. A consensus protocol that can take into account such issues is called *fault tolerant*, meaning that the protocol can tolerate (i.e., still works) when some nodes are faulty in this way.

The problem of achieving consensus in a distributed system has been known since the mid-eighties, and several fault tolerant protocols have been found and used. They essentially consist of specifying the types of messages that nodes send to each other and how those messages are used to update the data. For instance, one early and widely adopted solution for such cases is the communication protocol Paxos, and is used by many companies, including Google and Microsoft. At a high level, the Paxos algorithm is a three-step protocol that works as follows. Whenever a node, say, Alice, has new data that needs to be broadcast to the other nodes, she first contacts the other nodes proposing the data update. This is step 1, called the *proposal phase*. Since those other nodes may also receive proposals for data update from other nodes, each node must reply to Alice whether they will update their database according to Alice's proposal. This is step 2, called the *accept phase*. If Alice receives at least a certain pre-established number of positive answers to her proposal, then Alice will resend those nodes another message asking them to indeed update their database (and these latter reply to Alice confirming the update). This is the last step, called the *commit phase*. Alice does not wait for all nodes to accept in step 2, because some of them may be faulty. Consensus is obtained through the way the nodes respond to proposals like Alice's. When making a proposal Alice must add a number to her proposal. Nodes receiving various proposals will accept the proposal with the highest number and reject the others.

Note that there is a slight similarity with Bitcoin's consensus mechanism. When Bitcoin's blockchain forks, that is, when there are different versions of the blockchain, miners choose to mine on the longest blockchain. The highest number rule in Paxos' protocol is thus a little bit similar to the longest chain rule in Nakamoto's design. But we can also see a crucial difference between Paxos and Bitcoin. Under Paxos, there is a clear sequence of two-way communication between nodes. This contrasts with Bitcoin, where each node (miner) updates its blockchain and broadcasts new blocks of transactions to the network (if any) without waiting for some confirmation from the other miners. Under Paxos, Alice knows

how many positive answers to her proposals will be enough because the number of nodes in the system is known. As Bitcoin is permissionless, no one ever knows how many nodes there are, or even what exactly is the computational power involved in the system.

The presence of faulty nodes is in fact a minor problem. A more important one is to deal with the presence of malicious nodes. Consensus protocols that can account for the presence of malicious nodes are called *Byzantine Fault Tolerant*. The name Byzantine fault comes from a tale invented in 1982 by three computer scientists, Leslie Lamport, Robert Shostak, and Marshall Pease to describe the problem of communication between nodes in a network, which goes as follows. A number of Byzantine generals are contemplating attacking an enemy city but would only do so if all the generals agree. To make things worse, some generals may be treacherous. Complete and accurate communication is thus paramount. The problem is thus to find a communication protocol that can take into account what are called Byzantine faults, that is, possible errors due to a message of one of the generals not going through or a wrong message due to the presence of malicious nodes/generals. The Paxos protocol we have just outlined is not a Byzantine fault tolerant protocol, because it requires that all nodes trust the messages they receive, if they receive any. There is, however, a widely used Byzantine fault tolerant protocol, the Practical Byzantine Fault Tolerant (PBFT) protocol. This protocol is similar to Paxos in that nodes also wait to have a minimal number of other nodes agreeing to a proposal before implementing it. However, Paxos and PBFT differ in two important aspects: the amount of communication required and resilience. Paxos requires fewer messages being sent back and forth between nodes. That is, under Paxos a node proposing a data update (Alice in our example) communicates with several nodes and each of these nodes communicate with Alice only. In contrast, under PBFT each node contacted by Alice will communicate with other nodes as well. Reaching consensus thus may take more time under PBFT because more messages are exchanged. However, PBFT is more resilient, since the nodes contacted by Alice can still process the update even if Alice fails to respond, or sends different messages to different nodes. The choice between these two protocols (and their many variations) thus hinges on the tradeoff between speed vs. efficacy, and likelihood of faulty nodes vs. errors in communications between the nodes.

A key feature of most permissioned consensus protocols like Paxos or PBFT is that they rely on voting, that is, a sufficient number of nodes

must approve (and successfully communicate their approval) an update before it becomes effective. This notion of voting is what marks a crucial difference between permissioned and permissionless distributed systems. In permissioned distributed systems the number and the identity of participants is known because, by definition, each participant must be granted access before being part of the network. Voting thresholds can thus be fixed in advance by the consensus protocol. For instance, in a network with, say, 100 participants, it has been shown that Byzantine fault tolerant consensus protocols can be implemented as soon as there are 67 or more trusted nodes. Such voting mechanisms cannot work for permissionless systems. To see this, notice first that the voting threshold cannot be a percentage because the number of voters is never known in advance. So, when a node is receiving votes from other nodes it will be impossible to know whether the votes received are from, say, 10, 50, or 90% of the nodes. Voting thresholds must then be set in terms of the absolute numbers of nodes. This is a major problem because in this case the system would be susceptible to what computer scientists call Sybil attacks, that is, attacks consisting of creating a sufficiently large number of nodes.

To sum up, the key difference between permissioned and permissionless distributed systems is the degree of trust between participants. Permissionless blockchain protocols like Bitcoin's are well adapted for the extreme case where there is no trust altogether. For permissionless systems, consensus can be achieved through strong pecuniary incentives. Such strong incentives naturally arise when the platform is issuing coins that have monetary value. At the same time, recent academic work and real-life examples like Bitcoin Gold's attacks described in Chapter 4 have shown that security is more difficult to achieve when monetary incentives are too small relative to the cost of acquiring the majority of the computing power and the value of transactions that can be stolen. This is where permissioned blockchains, drawing on distributed systems may have an edge. Replacing monetary incentives with a certain degree of trust among participants opens the opportunity to using a wider range of consensus protocols.

Until now we focused on consistency of the data between the nodes in a distributed system. Yet, another concern that may arise is the reliability of the data. This concern arises in cases when the information in the blockchain relates to processes external to the system. A blockchain used to manage real estate transactions would be a system relying on external information. This is because while blockchain may reliably handle

transfers of value, the initial information about the real estate ownership, which is first created outside the blockchain, needs to be entered in the blockchain. We can easily imagine that if there would be no restriction on who can enter this initial information, it could be easily corrupted. Thus, reliability of the information on the blockchain hinges on trustworthy gatekeepers entering the information. We call it the *gateway problem*. In contrast, Bitcoin is free from the gateway problem, as all bitcoins are created directly on the blockchain, exist only on the blockchain and do not represent anything outside of it. Similarly, tokens set up on Ethereum are not subject to the gateway problem by themselves. However, as soon as they are linked to an object or information outside of the blockchain (e.g., an NFT representing an art piece), the gateway problem reappears. This is why NFTs issued in a completely permissionless system are not a reliable way of managing property rights of art, as we discussed in the previous chapter. The gateway problem is not confined to tokens. Smart contracts that rely on external data like our shipping or weather insurance examples we also discussed in the previous chapter are also subject to that problem. In the context of smart contract execution, it is often referred to as the *oracle problem*, related to the oracle risk discussed in that chapter. Permissioned systems have an advantage over permissionless one in solving the gateway problem, due to the ability of vetting validators and gatekeepers participating in the system, as well as removing the permission in case of misbehavior.

Permissionless and permissioned blockchains are thus optimal for different environments and purposes. Permissionless blockchains like Bitcoin are well suited to manage cryptocurrencies or tokens. By using the native coins to incentivize miners, a high level of security can be achieved even when there is no trust among participants. Absent the costly mining incentives to handle Byzantine faults, security can be obtained by making the platform permissioned. Thus, it is not surprising that many established organizations favor permissioned blockchains rather than trying to adapt Bitcoin's protocol for their purposes.

7.4 Enterprise Blockchain Solutions

Over the last several years there has been a number of projects aimed at offering blockchain systems specially designed for enterprise use, with a special interest on permissioned blockchains. One of those projects is Hyperledger, developed by a consortium comprising, among others,

technology companies (Cisco, IBM, Intel, etc.), financial institutions (JP Morgan, CME Group, Deutsche Börse Group, etc.), software companies (SAP), or academic institutions (Columbia University, UCLA, etc.). The project was started in 2015 by the Linux Foundation. The objective of the Hyperledger project is to define a general architecture, or framework for a blockchain system that can easily be tailored to any specific objective.

There are two important aspects of Hyperledger. Firstly, the framework developed by Hyperledger is aimed to be tailored specifically for enterprise use, improving the reliability and scalability compared to the existing cryptocurrency based blockchains. In particular, this implies that Hyperledger is a framework allowing for permissioned blockchains, and thus useful for networks of enterprises where there is a certain level of trust among participants. Secondly, the Hyperledger project does not consist of building a blockchain platform, but rather of establishing guidelines and standards for permissioned blockchains.

Several firms and organizations that are part of the Hyperledger project have used the Hyperledger framework to build blockchain solutions. One of the most advanced is IBM's Hyperledger Fabric, a platform especially adapted to manage databases shared by independent entities as it is the case for supply chain management. One key feature of Hyperledger Fabric (or simply Fabric as it is often referred to by IBM) is its modular implementation. Components or additional features can be added at will, making Fabric versatile and easily adaptable to a wide range of applications. Fabric, like other projects made under the umbrella of the Hyperledger project, includes a strong identity and confidentiality management. At its core Fabric is a permissioned blockchain that is maintained by entities called validators. Credentials to become validators are given by a third party, the permission issuer. For instance, a company that wants to monitor and control its supply chain (e.g., a retailer or a manufacturer) would be the permission user and the validators could be trusted partners along the supply chain (global carriers, banks, etc.). The role of validators is similar to that of miners for Bitcoin, that is, they first check the validity of transactions or data updates before storing them into the blockchain. Fabric can work with different consensus protocols. The most common is a modified version of PBFT, Sieve. The main difference between Sieve and PBFT is that the former can detect and remove, if needed, what computer scientists call non-deterministic requests, i.e., requests (like the execution of a smart contract) that do not always give the same output for a given input. Non-deterministic requests may also

be the result of ill-designed smart contracts (we have seen in the previous chapter that smart contracts are not necessarily bug-free). Avoiding such requests can be highly desirable when an organization wants to run audits.

Under Fabric the requests or data updates that the validators have to process come from users. Those entities can be, for instance, local farmers or producers, local carriers, etc. The rich structure of permission rights proposed by Fabric permits the permission issuer to give distinct rights to distinct users. For instance, a local farmer may have the right to submit a transaction to the validators related to its activity (e.g., the crops have been harvested) but not with respect to the activity of another user like a carrier. Access to the data may also be restricted. Still considering the case of supply chains, an intermediary may, for example, have access to the history of a shipment until it was handled by that intermediary but not after.

Supply chains are a natural area to use blockchain for two reasons. The first reason is that transactions or data presumably cannot be removed once they are added to the blockchain. The second property, too rarely mentioned but equally important, is that the order of transactions is maintained in the blockchain, as it is part of its data structure. This second property comes naturally in the case of Bitcoin: Alice cannot send bitcoins to Bob before she has received those bitcoins herself. Like monetary transactions, steps in supply chains have a chronological structure that make them natural candidates for blockchain applications.

One of the most famous uses of IBM's Fabric is Walmart's supply chain management. Launched in 2018, Walmart's initiative now requires its leafy green suppliers to use Walmart's blockchain. According to Walmart, traceability is now a matter of seconds instead of days as it used to be, a clear advantage in case of health issues like an E. Coli outbreak.[4] It is not too difficult to play the devil's advocate, arguing that the traceability standard sought by Walmart is essentially due to the digitization of all the steps in the supply chain from the producer to the shelves in the stores rather than due to the use of a blockchain. This inference is certainly sound, but it is also easy to argue that Fabric or other similar permissioned blockchain solution is an appropriate choice. Large retailers like Walmart rely on an extended and diverse supply chain, comprising local farmers, small and large carriers, health authorities, or logistic firms.

[4] https://corporate.walmart.com/newsroom/2018/09/24/in-wake-of-romaine-e-coli-scare-walmart-deploys-blockchain-to-track-leafy-greens.

In complex commercial relationships it is not desirable to share all data equally across all ecosystem participants. For instance, local farmers may not want competitors to have access to their transactions with Walmart's network. Fabric allows to differentiate permissions to access data (read and/or write). Also, the number and variety of parties involved in the supply chain require a high level of trust and reliability, properties that can be easily achieved with Hyperledger Fabric.

Hyperledger and its applications, with IBM's Fabric being one of them, is not the only project aimed at developing blockchain solutions for enterprise. Another well-established product is the blockchain Corda, developed by R3, a young, New York-based company. Like Hyperledger Fabric, Corda can be adapted to a variety of environments. It is implemented, either fully or running pilots, in several industries, for example, in banking (Sputa Banca DLT in Spain), aircraft part repair (Aerotrax), or Gold trading (Tradewind Markets). While Fabric is well suited for supply chain management, Corda's main targets are financial markets, banks, and trading platforms.

Overall, IBM or R3's blockchain solutions show that blockchain can prove valuable for enterprise use. That success is undoubtedly due to a stark departure from Bitcoin's design. No longer permissionless, enterprise blockchains are cleared from using strong incentive schemes for miners (validators) to maintain security and consensus. Those solutions are also able to mitigate the gateway problem, thus improving the reliability of the information in the system. A famous solution to that problem is the blockchain platform developed by the company Everledger for the diamond industry. Like Walmart's blockchain for green leafy products, the blockchain created by Everledger can follow diamonds throughout its production and commercialization. One key step in the diamond chain is when stones are cut and polished, giving to each gem a set of characteristics that make any diamond unique. A major part of the gateway problem is precisely solved during this step: scanning, modeling and cutting machines act like sensors, sending diamonds' characteristics directly to the blockchain.

7.5 Government Applications

Permissioned blockchains are not only useful for businesses; they may be also appealing to governments. Estonia's blockchain for ensuring integrity

of institutional data made headlines multiple times.[5] It is part of a larger effort to ensure data usability and safety in Estonia, which is sometimes called the *digital republic*.

Starting in 2001, the Estonian government focused on moving to the digital realm all the services that Estonians use in their daily life. All services or data such as taxes, medical records, property deeds, voting rights, to cite just a few, are not only entirely managed electronically but also made easily accessible. Some of the advantages of this project are cost reduction for the administration and efficiency gains. Data requests from banks, the administration or other services can be now addressed quickly and reliably. Elements of blockchain design, mostly focusing on linking data with hashes to achieve a tamper-evident ledger, started to be introduced in 2008.[6] The Estonian experience is in fact more a tale about the advantages of digitizing our administrative lives than a tale about distributed databases or blockchain. Of course, such a project would not have been possible without the distributed databases. The technology used by the Estonian government in this project is called X-Road, which was developed under the authority of the Estonian Ministry of Economy and Communication.

The digitization of most aspects of Estonians' lives is not really unique: most developed countries have reached a level of digitization comparable to that of Estonia. As of today, many people in most countries (can) pay their taxes online, inquiry about property deeds online, have access to medical records online, etc. But the case of Estonia differs from that of most other countries in an important aspect: all those electronic services or, more precisely, all the databases corresponding to these services are connected. Every resident of Estonia and every entity (organization, firm, etc.) has a unique identifier that is used by any database or service the person or the entity may interact with. Having a homogenous identifier across distinct databases has the great advantages of making the collection of data extremely efficient and cost saving. To see this, consider for instance the case of, say, Alice, who is applying for a loan to buy a house from Bob.

[5] See for instance https://www.newyorker.com/magazine/2017/12/18/estonia-the-digital-republic or https://qz.com/1535549/living-on-the-blockchain-is-a-game-changer-for-estonian-citizens/.

[6] https://e-estonia.com/wp-content/uploads/2020mar-nochanges-faq-a4-v03-blockchain-1-1.pdf.

In most countries the amount of information that Alice needs to provide is the same: she first needs to show that she is solvable. For that, Alice must show that she has an income (e.g., showing an employment contract and pay stubs), disclose her tax return, savings, and outstanding debts. All that information must be provided to the bank where Alice will submit her loan application. In parallel, Alice will have to make sure that Bob is indeed the owner of the house he is selling to her, have the list of all possible violations and outstanding fiscal debts attached to the house Bob is selling. Whether we consider Estonia or any other country, all this information comes from different sources. However, in most countries these sources are not related, meaning that Alice would have to contact each of those sources, put together all the information she retrieves and send it to the bank and lawyer or notary. From the point of view of the other parties (the bank, the brokers, etc.), the very fact that the information comes from Alice (and not directly from the services Alice contacted) may also raise suspicion, and thus can make the operation slightly riskier and costlier. In Estonia things are different. What Alice would have to do is to give temporary authorization to the bank to retrieve her tax return (or a part thereof), her labor contract, and pay stubs, etc. Such interconnected systems bring several advantages. First, the amount of work that Alice has to perform to collect all the necessary data becomes negligible. Second, all the parties involved in the transaction collect reliable information (directly from the source), thereby eliminating the risks of using inaccurate data and saving costs of verifications. Also, the data being handled electronically also means that it can be immediately used, thus avoiding potential errors in data entry. While they are interconnected, different systems communicate with each other in a secure and confidential way. The bank's system communicates with the Estonian tax authority and Alice's employer. Yet, confidentiality is ensured because Alice can select which type of data third parties like her bank have access to and can only do so after Alice has given her authorization.

The overall outcome of the Estonian project is widely positive. Thanks to X-Road, not only administrative routines have been simplified but there is also greater transparency. For instance, it is just a matter of a few clicks for any Estonian citizen to obtain financial information about their politicians. It is worth noting that, while technology has been key to the digitization of most of everyday life's aspects, nothing of this would have happened without any political will. Systems like X-Road are only here to manage data access, not to make the data accessible in the first place.

One telling sign of the success of the Estonian project is that other countries are starting to follow the Estonian example, with Finland working jointly with Estonia to connect their e-services through X-Road. The Estonian project provides another valuable message: the implementation of new technologies is more likely to be successful if the particular design is following the specific problem. The success of Estonia's project is partly due to the fact that the technology came second, as it was developed and implemented to address particular needs.

The digital information infrastructure in Estonia is an example of a distributed database, with use of blockchain for data integrity. However, its purpose is different from Bitcoin or Hyperledger Fabric blockchains, where different parties host copies of the database, thereby facing the problem of maintaining consensus across the various copies. The problem in Estonia is entirely different. The main purpose of the Estonian network is not to pool all available data into one giant dataset and distribute it (or a part thereof) across participants. Each vetted participant (government, banks, schools, etc.) keeps its own data and is not in charge of hosting and maintaining the data of other participants. For instance, the node corresponding to a hospital has the patients' data but not their banking accounts or tax returns. Of course, consensus is still an issue (most data servers are replicated for security reasons), but it is not the main objective of the project. The key purpose of Estonia's X-Road infrastructure is the interoperability of the data while maintaining confidentiality and security. The integrity of the data is supported by linking data with hashes to achieve a tamper-evident ledger, akin to Haber and Stornetta's "blockchain" of 1990's (before the term was even coined).[7] Thus, the blockchain use in Estonia's digital information structure is very different from Bitcoin's or Fabric's blockchain.

The Estonian project thus offers an interesting perspective on the debate around blockchain, and its popular understanding. The concept of blockchain was popularized with Bitcoin, and many people now define a blockchain as a distributed, append-only, permissionless database with a consensus mechanism like proof-of-work. Consequently, there is a debate whether permissioned blockchain can be called blockchains, for the simple reason of being permissioned. At the same time, the literal meaning of the word blockchain—blocks of data linked into a chain by hashes—would

[7] Haber and Stornetta (1990).

include the Haber and Stornetta's design used to timestamp digital documents. In fact, there is no commonly agreed upon definition of what a blockchain is: whether it has to be permissionless like Bitcoin or it can be permissioned like Hyperledger Fabric. In Bitcoin all the nodes have a full copy of the blockchain, which is not necessarily the case in Hyperledger Fabric. Bitcoin has a decentralized structure, while Haber and Stornetta's blockchain is centralized. Yet, what all these blockchains have in common is an important feature that not all distributed data systems have: the use of hash pointers to link data entries, which is essential to making the ledger tamper evident. It is perhaps the key feature that differentiates blockchain from other types of data systems. Yet, this feature may be implemented by different protocols depending on the purpose and environment. As the properties of the resulting blockchain crucially depend on the underlying protocol, it would be misleading to expect that any blockchain will have the same properties as Bitcoin's blockchain.

CHAPTER 8

Future Full of Possibilities

The twenty-first century's information technology has offered an advanced scope of programmability and digital security. This new environment nurtured the creation of a technology that provided an unprecedented flexibility for the design of new currencies and, through the blockchain technology, a groundbreaking tool for digital value transfer that goes way beyond currencies.

Nakamoto's extraordinary feat of building Bitcoin led to a surge of enthusiasm, bringing an outstanding number of innovations. Bitcoin was the first working solution to a long-standing problem but, as it is often the case in such situations, there was room for refinements. The proposals that emerged were aimed not only at improving several technical aspects but also at extending the range of possible use of cryptocurrencies. In spite of these efforts, cryptocurrencies are still far from mass adoption as means of payment. Instead, cryptocurrencies and tokens have quickly been adopted as investment assets. Except a small number of people who enjoy the relative privacy or speed (e.g., for international remittance) that cryptocurrencies offer, the increasing popularity of payment tools such as Paypal, Venmo, or Zelle suggests that innovation in *how* money is used may matter more than innovation in the form or type of money. With that perspective, Bitcoin and the early cryptocurrencies appear to be more a realization of some ideal rather than a solution to a real need.

H. Halaburda et al., *Beyond Bitcoin*,
https://doi.org/10.1007/978-3-030-88931-9_8

In retrospect, even crypto entrepreneurs seem to have become aware that cryptocurrencies are unlikely to replace fiat money. As we have seen, while the early cryptocurrencies aimed at being more efficient, less costly for the society or improving privacy, most recent projects focused more on providing a tangible service (file storage, voting mechanism, etc.). Getting access to some specific service gives an even stronger incentive for adoption, and thus the development of cryptocurrencies and tokens shifted, emphasizing now their purpose (what for) more than their nature (how). The emerging pattern indicates that crypto will end up more similar to platform-based currencies like Amazon coin, than to fiat currencies like the US dollar. Taking this perspective, one could expect to see many forms of currencies in use. As digital platforms develop, they will continue to experiment with digital currencies to better serve their business models. One might conclude that, in the end, cryptocurrencies and crypto-tokens do not bring much novelty compared to already existing platforms with their centrally managed digital currencies. But such a hasty conclusion neglects the decentralized aspect of crypto. By being free of any centralized, corporate control the ecosystem surrounding crypto may be more flexible, and thus may have greater chances to last. It also allows the creation of decentralized platforms using decentralized crypto. Having said that, we should keep in mind that the time of mass adoption of cryptocurrencies or tokens, *for a specific purpose* has not come yet either. Cryptocurrencies are still essentially perceived as investment assets.

The fact that cryptocurrencies did not become (yet?) what their inventors initially sought is not exceptional. After all, some inventions do not end up being used as intended, and some inventions are "accidental." Sometimes the technology behind some invention ends up having its own life. The microwave oven, which comes from the development of radar during World War II, is one out of many examples. Saying that the blockchain is now part of this club is in no way preposterous. Today, the word "blockchain" is without any doubt more ubiquitous and certainly inspires more respect than Bitcoin (e.g., drug trafficking is associated with Bitcoin, not blockchain). Satoshi Nakamoto may not be the inventor of the blockchain technologies at large, but has, as the designer of Bitcoin, unquestionably helped to unveil their potential.

Predicting what the future of blockchain will bring is difficult and hazardous, as much as it was the case for cryptocurrencies 5 or 7 years ago. There are, however, some similarities between the two: a strong enthusiasm for a complex and often poorly understood technology, which

is believed to solve important issues. There are also some notable differences between the trajectory of cryptocurrencies (and tokens) and that of blockchain technologies. While the development of cryptocurrencies or tokens was almost exclusively done by enthusiastic individuals or young startups, many relevant blockchain applications are produced by technology behemoths like IBM, Google, or Oracle.

For what it seems, such corporations did not see much interest in cryptocurrencies, whether it is with respect to the potential of cryptocurrencies or the lack of value to be harnessed from a business perspective. Applying blockchain technologies beyond cryptocurrencies seems to be more attractive for established players. Most of such "blockchain solutions," however, are permissioned. This suggests that the blockchain design likely to be adopted on a large scale in industry will be substantially different from the one envisioned by Nakamoto.

The enthusiasm around the benefits of blockchain technology may also have been misplaced. In some cases, most of the benefits may come from processes inspired by blockchain adoption than blockchain itself. For example, proponents argue that blockchain technology improves management of our data, but blockchain only works for digital data. Thus, to take advantage of these advocated improvements, the issue of digitization needs to be addressed first. Whether it is to track shipments or manage property rights of artworks, too often the solutions proposed implicitly assume a de facto digitization of our lives and activities, which in reality is still work in progress. Many industries are still paper based. While in some cases the lack of digitization is a bottleneck to blockchain adoption, in other cases, the enthusiasm for blockchain may provide an additional incentive driving a much-needed digitization. In other words, blockchain technologies can be seen as the spark that was needed to ignite a new push to digital society. It is in this sense that Bitcoin, or more generally blockchain, may be seen as a revolution.

References

Antonopoulos, Andreas (2014), *Mastering Bitcoin*, O'Reilly Media.

Bakos, Yannis and Hanna Halaburda (2019), "Funding New Ventures with Digital Tokens: Due Diligence and Token Tradability," available at https://ssrn.com/abstract=3335650.

Baloyan, Sergey (2021), "5 Most Expensive NFTs (Non Fungible Tokens) Ever Sold," *Hackernoon*, March 6, https://hackernoon.com/5-most-expensive-nfts-non-fungible-tokens-ever-sold-fd2t335j.

Barrera, Cathy and Hurder, Stephanie (2018), "Blockchain Upgrade as a Coordination Game," Available at SSRN: https://ssrn.com/abstract=3192208 or http://dx.doi.org/10.2139/ssrn.3192208.

Benetton, Matteo., Giovanni Compiani, and Adair Morse (2021), "When Cryptomining Comes to Town: High Electricity-Use Spillovers to the Local Economy," Available at SSRN: https://ssrn.com/abstract=3779720.

Blandin, Apolline and Gina C. Pieters, Yue Wu, Anton Dek, Thomas Eisermann, Damaris Njoki, and Sean Taylor (2020), "3rd Global Cryptoasset Benchmarking Study," Available at SSRN: https://ssrn.com/abstract=3700822.

Budish, E., P. Cramton, and J. Shim (2015), "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response," *The Quarterly Journal of Economics*, 130, pp. 1547–1621.

Caillaud, Bernard and Bruno Jullien (2001), "Competing Cybermediaries," *European Economic Review*, 45, pp. 797–808.

Caillaud, Bernard and Bruno Jullien (2003), "Chicken & Egg: Competition among Intermediation Service Providers," *RAND Journal of Economics*, 34, pp. 309–328.

H. Halaburda et al., *Beyond Bitcoin*,
https://doi.org/10.1007/978-3-030-88931-9

Casadesus-Masanell, Ramon and Feng Zhu (2010), "Strategies to Fight Ad-Sponsored Rivals," *Management Science*, 56, pp. 1484–1499.

Castronova, Edward (2014), *Wildcat Currency*, Yale University Press.

Comparette, T. Louis (1914), "Debasement of the Silver Coinage under the Emperor Nero." *The American Journal of Numismatics*, 47, pp. 1–11.

Einzig, Paul (1966), *Primitive Money: In Its Ethnological, Historical and Economical Aspects*, Pergamon.

Evans, David S. (2012), "Facebook Credits: Do Payments Firms Need to Worry," PYMNTS.com, February 28, http://www.pymnts.com/briefing-room/commerce-3-0/facebook-commerce-2/Facebook-Credits-Do-Payments-Firms-Need-to-Worry-2/.

Evans, David S. and Richard Schmalensee (2005), *Paying with Plastic*, MIT Press.

Farrell, Joseph and Garth Saloner (1985), "Standardization, Compatibility, and Innovation," *RAND Journal of Economics*, 16 (1), pp. 70–83.

Fergusson, Niall (2008), *The Ascent of Money*, The Penguin Press.

Fischer, Michael J., Nancy A. Lynch, and Michael S. Paterson (1985), "Impossibility of Distributed Consensus with One Faulty Process," *Journal of the ACM*, 32, pp. 374–382.

Fox, Armando and Eric Brewer (1999), "Harvest, yield, and scalable tolerant systems," *Proc. 7th Workshop Hot Topics in Operating Systems (HotOS 99)*, IEEE CS, pp. 174–178.

Fowler, Geoffrey A. and Juying Qin (2017), "QQ: China's New Coin of the Realm?" The Wall Street Journal, March 30. https://www.wsj.com/articles/SB11715196701146535l8.

Fung, Ben and Hanna Halaburda (2014), "Understanding Platform-Based Digital Currencies," *Bank of Canada Review*, April, pp. 12–20.

Gandal, Neil and Hanna Halaburda (2014), "Competition in the Cryptocurrency Market," WEIS Collected Volume.

Gandal, Neil and Hanna Halaburda (2016), "Can We Predict the Winner in a Market with Network Effects? Competition in Cryptocurrency Market," *Games*, 7, p. 16.

Gandal, Neil, J. T. Hamrick, Tyler Moore, and Tali Oberman (2018), "Price Manipulation in the Bitcoin Ecosystem," *Journal of Monetary Economics*, 95, pp. 86–96.

Gans, Joshua and Hanna Halaburda (2015), "Some Economics of Private Digital Currency," *Economic Analysis of the Digital Economy*, A. Goldfarb, S. Greenstein and C. Tucker (eds), The University of Chicago Press.

Gilbert, Seth and Nancy A. Lynch (2002), "Brewer's Conjecture and the Feasibility of Consistent, Available, Partition-Tolerant Web Services," *ACM SIGACT News*, 33, pp. 51–59.

Greenberg, Andy (2017), "Monero, the Drug Dealer's Cryptocurrency of Choice, Is on Fire," *Wired Magazine*. https://www.wired.com/2017/01/monero-drug-dealers-cryptocurrency-choice-fire/.

Haber, Stuart and W. Scott Stornetta (1990), "How to Time-Stamp a Digital Document," *Conference on the Theory and Application of Cryptography*, pp. 437–455. Springer, Berlin, Heidelberg.

Haeringer, Guillaume and Hanna Halaburda (2018). "Bitcoin: A Revolution?" *Economic Analysis of the Digital Revolution*, J. Ganuza and G. LLobet (eds), Funcas.

Harari, Yuval Noah (2014), *Sapiens: A Brief History of Humankind*.

Huberman, Gur, Jacob D. Leshno, and Ciamac Moallemi (2021), "Monopoly without a Monopolist: An Economic Analysis of the Bitcoin Payment System," *The Review of Economic Studies*.

Hof, Rob (2006), "Second Life's First Millionaire," *BusinessWeek*, November 26, http://www.businessweek.com/the_thread/techbeat/archives/2006/11/second_lifes_fi.html.

Jevons, W.S (1875), *Money and the Mechanism of Exchange*, Macmillan.

Katz, Michael L. and Carl Shapiro (1985), "Network Externalities, Competition, and Compatibility," *American Economic Review*, 75 (3), pp. 424–440.

Kiyotaki, Nobuhiro and Randall Wright (1989), "On Money as a Medium of Exchange," *Journal of Political Economy*, 97, pp. 927–954.

Lewis-Kraus, Gideon (2018), "Inside the Crypto World's Biggest Scandal," *Wired Magazine*, June 19.

Lorenz, Taylor (2021), "Digital Horses Are the Talk of the Crypto World," *New York Times*, May 1, https://www.nytimes.com/2021/05/01/style/zed-run-horse-racing.html.

Martin, Felix (2013), "Money: The Unauthorised Biography," Bodley Head.

McMillan, Robert (2013), "Ex-Googler Gives the World a Better Bitcoin," *Wired Magazine*, August, 30, https://www.wired.com/2013/08/litecoin/.

Morgenson, Gretchen (2021), "Some Locals Say a Bitcoin Mining Operation Is Ruining One of the Finger Lakes. Here's How," *NBC News*, https://www.nbcnews.com/science/environment/some-locals-say-bitcoin-mining-operation-ruining-one-finger-lakes-n1272938.

Murphy, Hannah (2021), "Monero emerges as crypto of choice for cybercriminals," *Financial Times*, June 22, https://www.ft.com/content/13fb66ed-b4e2-4f5f-926a-7d34dc40d8b6.

Nakamoto, Satoshi (2008), "Bitcoin: A Peer-to-Peer Electronic Cash System," https://bitcoin.org/bitcoin.pdf.

Pagliery, Jose (2014), *Bitcoin and the Future of Money*, Triumph Books.

Planet Money (2010), "A Chemist Explains Why Gold Beat Out Lithium, Osmium, Einsteinium…" November, 19. https://www.npr.org/sections/

money/2011/02/15/131430755/a-chemist-explains-why-gold-beat-out-lithium-osmium-einsteinium.

Rochet, J.-C. and J. Tirole (2002), "Cooperation Among Competitors: Some Economics of Payment Card Associations," *RAND Journal of Economics*, 33 (4), pp. 549–570.

Rochet, J.-C. and J. Tirole (2003), "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association*, 1 (4), pp. 990–1029.

Roth, Alvin E. (2015). *Who Gets What—And Why: The New Economics of Matchmaking and Market Design*, Houghton Mifflin Harcourt.

Saleh, Fahad. (2021), "Blockchain Without Waste: Proof-of-Stake," *The Review of Financial Studies*, 34, pp. 1156–1190.

Sargent, Thomas J. and Francois R. Velde (2002), *The Big Problem of Small Change*, Princeton University Press, 2002.

Sheck, Justin (2018), "Mackerel Economics in Prisons Leads to Appreciation for Oily Fillets," *Wall Street Journal*, October 2, http://www.wsj.com/articles/SB122290720439096481.

The Economist (2013), "Kill bill," March 16, http://www.economist.com/node/21573582/print.

Tschoegl, Adrian E. (2001), "Maria Theresa's Thaler: A Case of International Money," *Eastern Economic Journal*, 27 (4), pp. 445–464.

Vigna, Paul and Michael J. Casey (2015), *The Age of Cryptocurrency*, St. Martin's Press.

Vincent, Danny (2011), "China Used Prisoners in Lucrative Gaming Work," *The Guardian*, May 25, http://www.theguardian.com/world/2011/may/25/china-prisoners-internet-gaming-scam.

Vogelsteller, Fabian and Vitalik Buterin (2015), "EIP-20: Token Standard." https://eips.ethereum.org/EIPS/eip-20.

Weatherford, Jack (1997), *The History of Money*, Three Rivers Press.

Weber, Warren E. (2014), "The Efficiency of Private E-Money Like Systems: The U.S. Experience with State Bank Notes," Bank of Canada Working Paper No. 2014-15.

Yglesias, M. (2012), "Social Cash: Could Facebook Credits Ever Compete with Dollars and Euros?" *Slate*, February 29, http://www.slate.com/articles/business/cashless_society/2012/02/facebook_credits_how_the_social_network_s_currency_could_compete_with_dollars_and_euros_.html.

Index

H. Halaburda et al., *Beyond Bitcoin*,
https://doi.org/10.1007/978-3-030-88931-9

C