THE EMERGENCE OF BLOCKCHAIN:
INSTITUTIONS, OPEN SOURCE TECHNOLOGY AND THE
SOCIAL ORGANIZATION OF EARLY-STAGE DIGITAL TRANSFORMATION

Anwar Jason Windawi

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY

THE DEPARTMENT OF SOCIOLOGY

Advisers: Paul Starr and Viviana Zelizer

November 2021

Abstract

Blockchain technology has emerged from the shadow of Bitcoin to become a central component of 21st century digital transformation, and has become a focus of experimentation by actors ranging from the world's oldest central banks to startups spanning nearly every industry sector. This combination of novelty and seeming ubiquity means that blockchain technology provides an unusual opportunity to study digital transformation in its pre-settlement stage. This dissertation approaches digital transformation as a question of social organization, and asks how people organize to implement a novel technology when both the technology and the social organization around it are in flux. I explore this question of techno-social organization at the macro and meso levels using a combination of computational and qualitative digital methodologies. At the macro level, my analysis is based on a unique data set I have constructed of roughly 5,000 blockchain implementations and an interpretive data science approach I developed to classify these implementations in the absence of a "ground truth" pre-existing framework. I develop a three-level taxonomy of implementations that I then use to ask how an emergent general purpose technology interacts with higher-level, institutionalized social structures such as institutions and established economic domains. I find that the technology's generality of purpose is embodied by three clusters: programmable money, tokenization and computational infrastructure. I also find that its widespread adoption is driven by three patterns of reinvention whose differences result largely from variation in the alignment between the functions of the technology and the core institutions in the domains in which it is being applied. At the meso level I use comparative case studies and algorithmic ethnography to study how people organize to implement a new technology. My analysis at the meso level is divided into classic questions of coordination and control or governance. I find that coordination uses both established and 21st century organizational forms as component microstructures that are recombined in novel ways. In studying governance, I find that the institutionalization of open source processes has, together with the technology itself, enabled the emergence of a new, digital form of institution-building.

Acknowledgements

This dissertation would not have been possible without the support and insights of my committee. I'm particularly indebted to my co-chairs Paul Starr and Viviana Zelizer, each of whose work provided a template and inspiration that helped orient my own, and whose generous comments and guidance were indispensable. The institutional component of the dissertation evolved out of several years of dialogue with Paul Starr's work in the seminar room and as a scholar. Viviana Zelizer's pioneering work on money has been and will continue to be the touchstone for my thinking about sociology's important role in this new, digital landscape. I am also grateful to Mitch Duneier for his ongoing support during the dissertation.

My intellectual debts go well beyond those to my committee. I had the great privilege of having defense examiners whose work made my own possible: Karin Knorr Cetina and Brandon Stewart have each done pioneering work in economic sociology and computational social science on which I built for this project, and whose comments and questions during the defense will be fundamental to this project's future development. This work also grew out of engagement with the work of David Stark and Diane Vaughn, who continue to provide inspiration for its development. I am also deeply grateful to Marc-David Seidel and his colleagues at Blockchain@UBC including Victoria Lemieux, Chang Lu, Kam Phung and Chris Rowell, for a generative visit to UBC and ongoing conversations that have contributed to this project's development. In the same vein, this project has benefited from conversations with Miguel Centeno, Paul DiMaggio, Primavera De Filippi, Teppo Felin, James Grimmelmann, Ying-Ying Hsieh, Alex Preda, Carla Reyes, J.P. Vergne and Vlad Zamfir.

The actual work of writing this dissertation was only possible due to the support of two writing communities. The bulk of it was written during morning sessions of Writing with Randos, organized by our fearless leader Elizabeth Popp Berman, and including the most supportive case of characters a writer could hope for: Danna Agmon, Meredith Broussard, Derek Gottlieb, Amy Graves Monroe, Daniel Hirschman, Christina Ho, Elizabeth Mazzolini, Victor Ray, Deana Rohlinger, Louise Seamster, Jennifer Sessions, Adam Slez, Lisa Stampnitzky, Elly Truitt, Didem

Türkoğlu, LaDale Winling and Sarah Woulfin. Earlier versions were partially drafted in a peer writing group set up by Taylor Winfield, who remains an inspiration as a sociologist and as a human being.

This dissertation entailed a long hike through uncertainty that was only possible because of the unfailing support of family and friends. My sister Ashleigh Windawi was the best research assistant I could have hoped for in helping to build the data set at the core of the project. My sister Sarah Windawi and mother Elizabeth Scott were also always there with words of support that made a difference, as were my new extended family in Boston and Philadelphia: Betty Davidson, Jo Ellen Hillyer and Nancy Morris. Jim Brian Gordon and Katrin Smolka have been true friends and I am grateful to have met them in my graduate school travels. I am equally grateful for the support and friendship of friends and neighbors: Joanne Costanza, Gina Falsetto and Warren Brown, Lisa, Marc and Harrison Hugeuenin, Laura Strong and Eric Dutaud, and Liz Wilkinson and Jay Scheuerle.

Most important of all: my deepest thanks to my partner Joe Davidson. Drafting for this dissertation overlapped with the first four waves of COVID-19, as well as the loss of parents for both of us. We also lost our dog of 15 years as well as our annual refuge in Maine during that same period. Moving from this period of loss into whatever comes next is only possible because of you, and I am excited to see where we go next.

To the memory of Betty Davidson

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Webster Hall is a fixture of Manhattan's East Village, a concert venue and dance hall crowded several nights a week by rowdy partygoers who spill out onto the sidewalk for a smoke. It is a strange place to be at nine AM, and an even stranger place to be at nine AM on a Tuesday while doing the chicken dance with a roomful of suited bankers. The bankers and I were there on that April morning in 2017 to learn more about a relatively new technology called "blockchain," and how it might affect an even newer set of technologies referred to as "fintech," or consumer-oriented financial technology. The speaker on the stage - the managing partner of a venture capital firm known for its focus on financial technology - had led us all into this awkward moment as a way of embodying her assertion of the importance for "all of us to get uncomfortable" given the potential for disruption of the previously well-defined fields of banking and investment.

The awkwardness of the moment was also evocative of the ambiguity created by the cresting wave of hype about this new technology, which was in the process of emerging from shadow of the libertarian and Dark Web-evoking cryptocurrency Bitcoin. This hype was most apparent in the ecstatic tone of boosterism among many popular books that breathlessly extolled a new age of cryptocurrency and a "revolution" in which blockchain's digitized mechanisms for performing key institutional functions were "changing money, business and the world" (Tapscott and Tapscott

2016) and "challenging the global economic order" (Vigna and Casey 2016). The Economist provided a more sober but still striking set of claims in a 2015 cover article that christened blockchain "the trust machine" (Economist 2015). The article began with a brief reference to the cryptocurrency Bitcoin before pivoting to its main argument: although Bitcoin had become infamous for its use in illicit purchases on "dark" markets, the technology that made Bitcoin work as a decentralized system had far greater potential. In the article's words:

> *. . . most unfair of all is that bitcoin's shady image causes people to overlook the extraordinary potential of the "blockchain", the technology that underpins it. This innovation carries a significance stretching far beyond cryptocurrency. The blockchain lets people who have no particular confidence in each other collaborate without having to go through a neutral central authority. Simply put, it is a machine for creating trust.* (Economist 2015)

However faddish these early claims seem in retrospect, the intervening years have also brought a number of more sober efforts to grapple with the institutional, legal and operational implications of the technology (Berg, Davidson, and Potts 2019; De Filippi and Wright 2018; Iansiti and Lakhani 2017; Werbach 2018). These implications are playing out in experiments with this new technology across a diverse and growing range of institutionalized domains in the private, public and civil sectors that have been experimenting with this new technology over the past several years. Institutions ranging from the oldest central bank in the world to global commodity producers to the world's largest financial firms have stepped into this new world of technology whose roots lie in visions of obviating their very existence. At the same time, thousands of startups building on the technology have proliferated across all sectors of the economy, ranging from agriculture to green energy to venture capital, as well as in new domains enabled specifically by the technology. Their proliferation has been fueled by a combination of new forms of financing through "initial coin offerings" or ICOs, as well as the lowered barriers to entry inherent in the 21st century standard of building new ventures using open sourced, publicly available software code (Eghbal 2016).

In a broader sense, the current moment is one of large-scale change in which blockchain is only one of several potentially transformative and intersecting technologies that, collectively, are the engines of digital transformation. These technologies include artificial intelligence, machine learning, the internet of things (IOT), virtual reality and other technologies with the common characteristic of gathering and transforming enormous amounts of data through computation into some form of knowledge. This move into a global economy in which data is "the oil of the digital era" (Economist 2017) signals a move into a near-future in which the internet - now decades old - is no longer the source of innovation, but rather a taken-for-granted background element in a longer-term shift away from production of physical goods and services to one in which computation is a primary mode of economic action.

Existing literature describes such unsettled periods as ones of techno-economic "ferment" in which new ideologies and new modes of organizing social action emerge and compete with entrenched incumbents (Perez 2003; Starr 2019; Tushman and Rosenkopf 1992). Settlements that follow such periods are defined by a combination of top-down constitutive choices (Starr 2005), institutional pressures (DiMaggio and Powell 1983; Starr 1982), and emergent patterns created by local decisions of individual and organizational actors whose heterogeneous behavior collectively determines the path of adoption and adaptation (Hannan and Freeman 1977). Organizational actors are particularly important to this process given their central role in the development, implementation and refinement of new technologies (Bowker 1994; Star and Griesemer 1989; Tushman and Anderson 1986). While there are rich literatures on institutions as stable and self-reproducing entities, less is known about their emergence and early development prior to stabilization, a period when both new technology and the organizational forms that implement it are "between diffusion and institutionalization" (Colyvas and Jonsson 2011; Katz and Gartner 1988).

Blockchain technology offers an unusual opportunity to study the coevolutionary relationship between technology and social organization when both are in flux. Rather than focusing purely on the technology, or on its effects on a single form of organization, I instead use this opportunity to focus on the intersection of technology and social organization more broadly by zeroing in on the process of implementing the technology. This allows me to ask: *How do people organize to implement a novel technology when both the technology and the social organization around it are in flux?* The nature of the technology itself is an invitation to push this question further to ask closely related questions at the meso and macro levels. First, how do people organize to implement a new technology when both the technology and the organizational templates available to entrepreneurs are defined by their opposition to centralization? At the macro level, how does an emergent general purpose technology (GPT, a term I define further below) interact with higher-level, institutionalized social structures such as institutions and established economic domains? Put another way, what is the social organization of an emergent general purpose technology?

Sociology's roots as a discipline lie in an earlier profound period of transformation, one in which longstanding forms of social organization were reshaped by, and in turn affected the development of new technologies (Durkheim 1997; Marx 1992; Weber 1978). A small but growing literature in the discipline is bringing a sociological lens to the current digital transition (Daniels et al. 2017; Lupton 2014; Sassen 1999, 2002) and the ways in which digital mechanisms can both upend some institutional frameworks while entrenching others (Fourcade and Healy 2017; McMillan Cottom 2017). While much of this literature focuses on existing platforms, the breadth of blockchain technology's applications provides an opportunity both to extend the insights of digital sociology and to revisit the core concerns of organizational, economic and institutional sociology in this new light.

As scholars of new forms of organization have pointed out (Puranam, Alexy, and Reitzig 2014), what appears to be novel in one context has often been thoroughly studied in others. This is especially the case when novel recombinations incorporate components that have been institutionalized in their own right (Meyer and Rowan 1977; Sabel and Zeitlin 1997). My goal in this chapter is to weave together recent as well as foundational work in the social sciences to develop an overarching framework within which I develop the remainder of the dissertation.

1.1 What is Blockchain? An Introduction by Analogy

Any answer to the seemingly simple question “What is blockchain?” is likely to include some combination of a gesture to its origins in Bitcoin and perhaps a reference to its threefold mantra of being *decentralized*, *trustless* and *immutable*, though exactly what these terms mean is often unclear (Weick 1990). The approach I take in describing the technology in Chapter 2 and throughout the remainder of the dissertation is to deal with its complexity directly, though in specific contexts for each chapter. For the moment, I introduce it more gently by analogy with another recent attempt at decentralized technology.

For all its mathematical and technical complexity, blockchain technology is at its core an evolution of the peer-to-peer (P2P) network paradigm, in which nodes of a network operate autonomously toward a shared end rather than being centrally managed. The relationship between P2P networks and blockchains is clarified by the example of Napster, an earlier innovation that made similar claims to decentralization. As a peer-to-peer file sharing network, Napster connected users’ computers using software, creating a supply of music that aggregated the holdings of each individual peer/user into a single, global list. As a consumer-oriented service, Napster emphasized ease of use and downplayed the technical “backstage” (Goffman 1959) in favor of a single user interface that emphasized a list of available files as well as the files the user was sharing on the network. Napster’s ultimate demise as a free service resulted in part from its creators’ decision - one that reflected technology constraints of the time - to index the data stored across participants’ computers in a single, centralized server that provided participants with that single view of the available data.

Unfortunately for Napster, this centralization also provided the music industry with a focused site for legal action. In the parlance of distributed systems, this is known as having a single point of failure. The central technological innovation of blockchains is their ability to function analogously to Napster by linking webs of devices performing some kind of computational work, but without the need for any centralized coordination. By decentralizing the network, blockchain technology creates a distributed system that aggregates the collective computing, storage and other capacity of a set of hardware, and does so without necessitating that users trust a single coordinating entity. This coordination is instead performed through protocols and software, which encode a set of incentives that are made more effective through the use of dedicated coins or tokens that operate within the system.

This combination is evident in the case of Audius (Rumburg, Sethi, and Nagaraj 2019), a blockchain-based music project that reflects both the original intent of Napster as well as subsequent advances in streaming technology such as Apple Music and Spotify. Like Napster, Audius seeks to provide a means for individuals to access and interact with music. Like the for-profit services, Audius has the goal of compensating holders of rights to the music through fees to users. Where Audius differs is in its use of blockchain and its component technologies to establish finely grained roles for all the parties who interact with it, and to share the fee revenues from those interactions more broadly and transparently than has been the case with streaming services. In doing so, Audius effectively inverts Napster's approach by opening the entire technology to participants and by emphasizing the articulation of a set of roles available to participants that comprise the network, rather than a simple user interface.

On Audius, music publishers and artists can register their music and property rights within the network, at which point the music becomes available for fee-based streaming, for which they are compensated. End users pay for access to music but can also earn fees for providing reviews and contributing other services. More generally, the platform replaces the industry's notoriously dense layers of intermediaries between artists and audiences with a set of protocols that establish roles, as well as rules and rewards/sanctions for each role based on its functions. In addition to the obvious ones of producing and consuming music, these functions include providing servers for content storage/serving, indexing content across servers for search, curation of content through reviews, and arbitration of disputes, each of which can theoretically be performed by anyone with the correct equipment and connectivity. The protocol defines the compensation for each of these roles as varying forms of network transaction fees.

While user fees charged for accessing content remain constant, the Audius protocol introduces an additional variable fee element for other network participants that is increasingly common among blockchain projects: staking. Staking is posting a bond or earnest money in a project's native coin or token in order to earn the *right* to provide a service within the project, and is taken as a demonstration of commitment or "skin in the game." Staking appears in multiple forms in blockchain projects that range from incentives attached to particular roles (as in the case of Audius) to forming the basis for the crucial process of nodes reaching agreement on the state of the blockchain (as in the case of "proof-of-stake" successors to Bitcoin's (Nakamoto 2008) intentionally wasteful "proof-of-work" consensus mechanism). Audius follows the broad outlines of many of these approaches by specifying that participants in its network can earn stake-based fees. The greater the amount of Audius tokens an individual stakes, the greater the likelihood that the network will select that individual to provide a service, leading in game theoretic terms to a higher expected payoff. Staking also adds a form of symmetry to payouts in that misbehavior by those who stake tokens can be punished by "slashing" or appropriation of some proportion of the amount staked.

1.2 Studying Technological Change in Progress: Analytical Challenges and Approach

A dissertation is a challenging endeavor under any circumstances, and is even moreso when its topic is rapidly evolving. I began this project in 2017, a period in which literature on the topic outside of computer science was sparse and the technology itself was constantly changing. Although the literature has since diversified significantly, the technology itself remains in profound flux, as does the range of its application. The fact that my topic is the social organization of these applications means that this expansion becomes the central challenge of developing a research approach. There are no classic chapters in social science methods textbooks to refer to in studying such a phenomenon as it happens. Nor is there a satisfactory single research site given the very nature of dispersed systems that rely on both global standards and local specialization. They inherently lack an "absolute center from which control and standards flow" (Star and Ruhleder 1996, p. 211) in dispersed systems that rely on both global standards and local specialization.

These challenges would seem to argue against taking such a new technology as a research subject for a dissertation. Yet it is precisely this combination of unsettled times (Swidler 1986) and technological change that defines digital transformation as a social phenomenon. It also provides an unusual opportunity to engage directly with the phenomenon as a sociologist, or in Donna Haraway's more evocative terms, to "stay[] with the trouble" (Haraway 2016). Uncertainty comes with large-scale transitions. Evolutionary biologist François Jacob describes this kind of uncertainty as inherent to both evolution and the progress of science (Jacob 1982). In Jacob's framing, this uncertainty is the result of an ongoing encounter between the possible that we imagine and the actual we can measure:

> Whether in a social group or in an individual, human life always involves a continuous dialogue between the possible and the actual. A subtle mixture of belief, knowledge, and imagination builds before us an ever-changing picture of the possible. It is on this image that we mold our desires and fears. It is to this possible that we adjust our behavior and actions. In a way, such human activities as politics, art, and science can be viewed as particular ways of conducting this dialogue between the possible and the actual, each one with its own rules (Jacob 1982, p. vii).

While there are few methodological guidelines for this kind of work, sociology does provide guideposts. In his pathbreaking study of technology adoption in hospitals, Barley (1990) identified a set of obstacles to the effective study of the relationship between new technology and organizational structures, including ambiguous terminology, "distant knowledge" accrued from historical abstractions, a lack of compelling linkages across social levels (see also Knorr Cetina 1981), and an emphasis on social structure at the expense of social action.

My analytical framework takes two notions from academic work at the intersection of economic sociology and science and technology studies (STS) that address several of the issues raised by Barley. The first is Latour's (1987) dictum to study new technology by following the actors and the technology inductively as they travel and evolve. In this framework, the movement of a new technology is less a process of undifferentiated diffusion and more one of translation (Callon 1984; Latour 1990) and reinvention (Rogers 1995) through the adaptation and development of context-specific versions of the technology. My commitment to following the actors, when coupled with the breadth of implementations of the technology I study, leads me to look as broadly as possible across implementations of the technology in pursuit of generalizable insights, rather than to follow the traditional STS path of single case studies.

The second notion I take from economic sociology and STS is to attend to materiality by avoiding artificial distinctions between between "technology" and "the social" (Knorr Cetina 1997; Latour 1987, 2005; Orlikowski and Scott 2008). This stance leads me to focus on the mutually constitutive relationship between the technology and the social organization involved in its implementation. This relationship is among the foundational concerns in 20th century organizational sociology (Lawrence and Lorsch 1967; Thompson 1967; Woodward 1958) and has found more recent applications as, respectively, the mirroring hypothesis in organizational literature (Colfer and Baldwin 2016) and Conway's law (Conway 1968) in computer science, each of which theorizes a direct relationship between the nature of technology and organization. I attempt to address this by treating social organization as an assemblage (DeLanda 2006) of social and technological components (similar in spirit to Latour's technogram and sociogram) that jointly comprise the implementation of a specific technology. The jointness of this relationship is central to my research given that both the technology and the forms of social organization involved in its implementation are evolving together.

Studying an emergent and potentially general-purpose technology adds a closely related set of challenges that add to those identified by Barley - complexity, heterogeneity and emergence. I shaped the analytical framework described in the remainder of this section around their articulation as well as my approach to addressing them.

1.2.1 Heterogeneity

Focusing on the social organization of implementations does narrow the space of inquiry somewhat, but the sheer breadth of those implementations still remains daunting. The applications span economic sectors (e.g. healthcare) and institutional domains (e.g. money) as well as functions that work across multiple sectors and domains (e.g. escrow, cloud computing). Moreover, the implementations themselves often blend aspects of multiple organizational and institutional forms, including polities, economies, corporate forms (both profit-oriented and non-profit), networks, markets and algorithms-based coordination and governance. The standard approach to heterogeneity in social science often amounts to reducing it to variation along a small number of dimensions. That approach is better suited to settled situations in which such variables are well established than to emergent and general purpose technologies in their early period of ferment.[1] In reality, complex social, technological and ecological systems have so many potential interaction effects that ecologist Hardin's warning that "we can never do merely one thing" (Hardin 1985) is particularly apt.

My generalized organizational framework of identifying components and the means of integration, when applied more broadly, provides a path through this heterogeneity. I capitalize on this approach by looking more specifically for *complementarities* as a primary basis for integration. Complementarities are a relational form that, at their most general, exist in situations of mutual interest in which a benefit for actor A also helps actor B. Although this concept is most closely associated with economics, where it broadly defines a production relationship based on prices and output, I draw on three more targeted and relevant bodies of work in which the term is defined more usefully.[2]

The first of these one of the definitional aspects of General Purpose Technologies, or "GPTs."

[1] Even then, critics argue compellingly that this reduction requires the analyst to ignore important information (see e.g. Brandtner 2017; Glynn, Barr, and Dacin 2000).

[2] Purely economic complementarities are often called "Hicksian" complementarities, and exist in net form when a reduction in the price of A leads to an increase in demand for B without a change in output, and in gross form when the same relationship leads to an increase in output. Another economic form is the "Edgeworth" complementarity, in which an increase in demand for A leads to an increase in demand for B.

These technologies are widely studied in historical economics as relatively rare instances of innovations that find wide use in across the economy due to innovational or technological complementarities (Bekar, Carlaw, and Lipsey 2018; Bresnahan and Trajtenberg 1995; Carlaw and Lipsey 2002). The steam engine and the Internet are classic examples. Complementarities arise from the potential for productivity gains from applying the GPT in unrelated "application sectors," creating incentives for both the originators of the technology and those who might benefit from it to invest in developing it further. These incentives are framed in this literature in terms of spillovers of benefits to external parties (such as network effects) and the resulting tensions around governance and investment that these spillovers create. Complementarities are central to my analysis of the overall ecology of blockchain implementations in Chapters 3-5 because they serve as a mechanism for the widespread adoption of the technology.

The GPT literature provides an additional insight for my analysis in its treatment of the internal heterogeneity of GPTs themselves. Such technologies are described as arriving in clusters that nest within one another such that a single technology may appear as both a component of a larger cluster of GPTs and as a GPT in its own right (Bekar, Carlaw, and Lipsey 2018; Carlaw and Lipsey 2002). For example, the Internet was both enabled by the spread of personal computers (another GPT), which in turn was enabled by (and drove significant investments in) the computer chip. This insight supports my treatment of "blockchain technology" throughout this dissertation as a term of convenience that can more precisely be specified as a cluster of technologies (e.g. the ledger, the P2P network, the smart contract, the token, etc), various components of which find use in various application domains.

Political economists have similarly focused on clusters of institutional arrangements in their macro-historical studies of the determinants of national economic performance (Aoki 1994; Hall and Soskice 2001). For example, Aoki's study of the relationship between corporate governance and employment relations in Japan found a specific kind of complementarity: a set of institutional arrangements that increased the country's economic performance only when they appeared together. Scholars of institutional complementarities focus on institutionalized domains (e.g.

industrial relations, corporate governance, etc.) in which these clusters of institutions appear.

Finally, organizationally, the notion of complementarity is central to the definition of ecosystems, in which underlying platforms - whether technical and/or economic - provide the basis for elaboration by "complementors" that extend the functionality of the platform and provide the basis for the collective capture of value. For example, the Apple App Store is an economic platform on which app developers can build their own offerings, in the process creating gains for both themselves and Apple. The power dynamics in this relationship are complex given the role of Apple in controlling the terms of exchange and pricing together with its reliance on app developers to collectively provide the products that make the platform economically valuable. This asymmetry is present, if not as nakedly, in most technological platforms to the extent that any changes in the technical specifications and protocols that define those platforms require corresponding shifts in the specifications of the apps and other technologies that rely on them. The extent of mutual coordination and negotiation over such changes in specifications and protocols is thus of primary importance, and leads to my emphasis on protocols and governance in Chapter 7.

1.2.2 Complexity

The extraordinary complexity of blockchain technology requires analytical decisions about how best to engage with it meaningfully while not straying too far into technical detail. Analytically, a commitment to following the actors, and focusing on the sociomaterial organization they enact, naturally leads me to focus on "technology in use" (Edgerton 2007), or "implementations" as I call them, rather than hypothetical "use cases" or the study of a single application such as Bitcoin. I attempt to capture those implementations, wherever they occur. This provides me with an unusually rich perspective on the breadth of uses and a sufficient base from which to draw preliminary conclusions. Focusing on technology in use also brackets questions of novelty and significance, which can only be determined in the future.

Two interlocking forms of sociotechnical organization are foundational for implementations of blockchain technology. The first of these is the *distributed blockchain project*, which is the sociotechnical form involved in creating and launching individual implementations of the technology. I define these as meta-organizational configurations, which is to say they incorporate both individuals and other organizations as members, and do so following broad architectural guidelines rather than having a single canonical form (Gulati, Puranam, and Tushman 2012; Meyer, Tsui, and Hinings 1993). The configuration I develop in Chapter 6 encompasses many of the core components standard to definitions of emergent organization (Katz and Gartner 1988). But, rather than being organized through purely market or hierarchical relations, the essential functional components and resources are jointly created with independent, external communities,. This pattern reflects the norms of the open source and other platform-based communities out of which Bitcoin, Ethereum and the rest of the ecology has grown, as well as broader trends toward the elision of organizational boundaries in order to capitalize on external but aligned communities (O'Mahony and Lakhani 2011; Parker, Van Alstyne, and Jiang 2017) to build complementary applications. It also reflects broader practices of organizing new efforts around project teams that work within, across or (in the case of open source code) as the core of new ventures (Adler and Heckscher 2007; Obstfeld 2017).

The second basis of sociotechnical organization in the overall ecology is the *blockchain protocol*. Protocols function differently in blockchain than in settled contexts. Blockchain is based on disciplines such as cryptography and distributed systems design that are embedded in traditions of open science. In these fields there is a constant search for ways to "break" existing systems and develop more secure alternatives. As a result, there are few established, widely adopted blockchain protocols in the institutionalized sense. Blockchain protocols instead represent self-identified attempts to propose new ways of achieving specific ends that range from establishing secure blockchain networks to token-based lending. They do so by proposing novel combinations of existing protocols, algorithms and cryptographic elements. More scientific protocols (typically developed by or in partnership with computer scientists) introduce new elements in addition to recombining existing components. Consequently, I treat the *blockchain protocol* as the algorithmic counterpart to the meta-organizational *blockchain project*.

Blockchain protocols also provide a basis for the integration of projects into larger networks or ecosystems. The wholesale adoption of the platform architecture by blockchain projects (as I describe in Chapter 6) has in turn necessitated the creation of publicly communicated protocols for use by third-party developers and entrepreneurs to build on and extend a given project's technology and functionality. Scholars of information systems call such tools "boundary resources" (Eaton et al. 2015; Ghazawneh and Henfridsson 2013) and note the tension in their role of attempting to stabilize some level of control for the platform provider who "owns" and creates these resources while also maintaining open participation by external actors. This tension makes protocols the primary locus of governance among and between blockchain projects, leading me to focus on protocol-oriented governance rather than resource coordination in Chapter 7.

1.2.3 Emergence

The temporal aspect of my research subject is perhaps the greatest challenge, in that there is no well-established way to study a new phenomenon as it takes place, nor can there be any realistic expectation of knowing from the outset what trajectory the research will take. My work here contributes to broader conversations about emergence (Seidel and Greve 2017, Powell, Packalen, and Whittington 2012; Powell et al. 2017) by focusing on the forms of social action that drive the development of new alternatives during emergence. In particular, I focus on entrepreneurship, which I define broadly as the assembly, combination and coordination of resources in pursuit of a collective goal. This definition allows for entrepreneurs in large organizations as well as startups, in either the public or private sector, and - crucially, given the importance of open source development practices - outside of traditional organizations. This multi-actor definition avoids the trap of linking entrepreneurship with the heroic Schumpeterian individual (Swedberg 2000).

I also treat entrepreneurship as a form of situated action (Vaughan 1998, 2014), though since I am looking across multiple organizations rather than at a single case, I define it as a form of action situated in multiple contexts. This approach is closest in spirit to Emirbayer and Mische's "relational pragmatics" in specifying that action is both situated both relationally and temporally (Emirbayer and Mische 1998). In what follows, I define and explore three modes of situated action that further articulate the components of relational pragmatics as relevant to my dissertation.

First, entrepreneurship is *culturally situated* in its reliance on existing frameworks of meaning. These frameworks accrue with particular force for entrepreneurs when they are formalized as identities, categories and other forms of classification on the basis of which audiences judge the legitimacy of the entrepreneur's claims (Zuckerman 1999; 2017). When both the entrepreneur's project and the market it is involved in creating are new, the focus of communication shifts from alignment with institutionalized categories to a mutual process of sensemaking (Weick 1995) and sensegiving about emergent categories (Durand and Khaire 2017; Kennedy 2008; Navis and Glynn 2010). Language and narratives are central to this process (Lounsbury and Glynn 2001; Wry, Lounsbury, and Glynn 2011). The central role of narratives, stories and rhetoric is evident elsewhere in this and other literatures, whether the cases involve new organizational forms (Ruef 2000), emergent industries (Granqvist, Grodal, and Woolley 2012; Grodal, Gotsopoulos, and Suarez 2015), new financial practices (Zelizer 1979) or new genres of art (Khaire and Wadhwani 2010).

Second, entrepreneurship as a form of agency is *temporally situated*, with a dual orientation to both the present and the future (Emirbayer and Mische 1998). Situating entrepreneurs temporally in the present provides a mechanism for the imprinting of existing institutionalized forms on new organizations; put another way, Schumpeter's entrepreneur can only exist if there are components available for her recombine. Powell and Sandholtz (Powell and Sandholtz 2012) follow Stinchcombe (1965) in noting that such forms reflect the contextual conditions at the time of their founding, a claim that Johnson (2007) extends to a broader argument that the components out of which organizations are built are also time-dependent. Martin (2009) notes that the disruption of institutions and fields typically leads to the construction of new structures from elements that remain; put another way, the environment may well be littered with organizational building blocks (Meyer and Rowan 1977) even when the structures they formerly comprised no longer predominate. Powell and Sandholtz synthesize these arguments into a broader claim that the study of organizational genesis should be conducted as a "sociology of compounds" that examines new forms in light of the components out of which they are built as well as the sources of those components. I approach such recombination through the actions of entrepreneurs creating these new compounds.

Entrepreneurship is also crucially situated in a sense of the future. The entrepreneur "is the analytically distinguished social type who has the ability to take a reflective position towards institutionalized practices and can *envision* alternative modes of getting things done" (p. 786, emphasis in original). (Beckert 1999). Developing and expressing visions of the future through narratives is a necessary condition for entrepreneurs to be able to gather resources and sustain new ventures amid uncertainty (Beckert 2016; March 1995). Entrepreneurship is a form of experimentation based on theories of the future (Felin and Zenger 2017), and on the imagined affordances of new technologies (Nagy and Neff 2015, van Lente and Rip 1998).

Finally, I argue that entrepreneurship is *socio-materially embedded* in communities that provide crucial resources (Ostrom 2005; Rao 1998). Ostrom's framework situates actors in resource environments jointly determined by the expectations and history of the community, the nature of the resource and the institutionalized rules. This framework does the subtle work of foregrounding the resource in question, which is not always obvious in the case of digital communities. Its emphasis on identifying roles also calls for attention to the composition of these communities, as well as these communities' goals and motivations for participating. Given the need for the designers of blockchain projects to assemble resources from different communities, I expand Ostrom's framework to situate these entrepreneurs within these multiple communities (O'Mahony and Lakhani 2011; Rao 1998), and examine the methods they use to coordinate resources from each.

1.3 Research Design and Findings

Managing this combination of heterogeneity, complexity and emergence necessitated both the theoretical moves I describe above as well as a flexible and integrative approach to methodology and data collection. I developed this approach in multiple stages that varied by levels of social organization, and drew from across branches of sociology and, where necessary, other traditions. As a first approximation, my approach was to use a combination of existing literature and general characteristics of the technology to identify the most relevant analytical variables as well as computational and qualitative methods, and then to apply those methods. Doing so led to multilevel constructs that varied from classification taxonomies at the macrosocial level to layered organizational and governance frameworks at the meso level, but shared a thematic commitment to reflecting the complexity of the terrain without allowing it to dominate the analysis.

This ecumenical approach underwrote the findings that mark this project's contributions. Developing an interpretive data science methodology for large-scale, text-based classification in the absence of underlying categories, and applying this method to my unique data set, allowed me to establish a taxonomy of implementations of the technology that reflected real-world experimentation across the widest possible breadth of domains. This analysis also provided the rich computational data I used to identify the general purpose aspects of the technology, namely programmable or purpose-built money, distributed computational infrastructure and the creation of digital assets with value. I then combine these findings with the custom taxonomy I developed to identify patterns in the sprawling diffusion of the technology. The relationship between technology and its diffusion, I find, is a function of the alignment between the general purpose aspects of the technology use and the core institutions of the domain in which it is being adopted. The three patterns of reinvention I identify in my data and analysis identify three different trajectories of reinvention driven by variation in technological-institutional alignment.

My analysis at the meso level was grounded in these insights but shifted to qualitative and comparative case methods in pursuit of more generalizable findings regarding structures and mechanisms. I broadly structured this analysis around the two core concerns of organizational management: coordination and control. My inquiry into coordination uses comparative cases and finds that blockchain implementation projects incorporate multiple forms of distributed organization that are common building blocks for 21st century technology entrepreneurs but not yet widely studied in economic and organizational sociology, and are largely absent from any literature when considered as components of larger assemblages. I identify the component resources of these structures (the project core, codebase, computation, capital and platform) and the processes that projects use to co-construct these resources through exchange with external communities. In contrast to public claims that blockchain technology eliminates the need for traditional hierarchy, I also identify the crucial role of formal organizations in the design and enactment of mechanisms for these relational processes.

Rather than inductively assembling an organizational construct, as I do in the prior chapter, my analysis of control and governance begins with the construction of a multilevel, multivariable analytical framework grounded in the work of Ostrom (2005) and Starr (2019), and then uses algorithmic ethnography to populate the framework's levels and variables. Applying this framework to blockchain protocols, I identify and analyze three aspects of governance in the ecology: governance by protocols, governance of protocols, and constitutional transitions via protocols. My primary finding is the central importance of open source processes to the construction of digital institutions, a process I call open source institution-building.

1.4 Dissertation Outline

Chapter 2 by provides two complementary perspectives on blockchain technology. The first perspective is grounded in the technology itself and defines its components before describing how they are often combined in practice. The second is more historical and organizational and traces the evolution of the technology from its origins in Bitcoin to the most recent developments at the time of this writing. Each of the three empirical chapters that follow incorporate limited descriptions of the technology as appropriate to their intended function as standalone papers. But this chapter provides the most general treatment meant to support the entire dissertation.

Chapters 3 to 5 take on the question of the technology's early diffusion from the perspective of macrosocial organization. Chapter 3 motivates and articulates the methodological approach I develop to answer this question, which I ground in both the precedents and gaps in the sociological literature on categorization and the historical economics literature on general purpose technologies. I draw on an original data set I have constructed of more than 5, 500 project documents on which I use a combination of computational and interpretive approaches to develop the insights into blockchain's nature as a general purpose technology in Chapter 4, as well as its diffusion across other domains in Chapter 5.

The following two chapters shift from the macro to the meso level of sociotechnical organization and seek more generalizable answers to the question of how a 21st century GPT enters the world during its period of ferment. Chapter 6 focuses on the ongoing coordination of resources by inverted organizations in which nearly every key resource is outside of the bounds of traditional organization.

Chapter 7 shifts from the challenge of coordination to that of control, and focuses on the relationship between protocols and governance. Protocols are typically the site of the most intensive governance activity in blockchain projects, and thus provide a point of entry for a more detailed analysis of the sociotechnical aspects of governance in this new setting.

Chapter 8 concludes with a discussion of findings, and implications for future work.

Chapter 2

Blockchain as a Potential General Purpose Technology

Digital transformation as a sociological phenomenon poses a paradox. While it has been studied widely in the context of specific technologies such as algorithms and their effects on various social settings, it nonetheless lacks not only a unified definition, but a sense of its scope. This lack of definition points to the need to give structure and substance to the notion of "transformation," giving a sense of what is being transformed, and how. While the former question of "what" implies an emphasis on what already exists (and is thus a potential site of change), the question of "how" also opens the generative possibility of new domains of knowledge and technology.

In practice, the "how" of digital transformation is grounded in a set of enabling technologies (Teece 2018) that include artificial intelligence (AI), machine learning (ML), blockchain, cloud computing, and the Internet of Things (IoT). These technologies, scholars have claimed, have the attributes of a specific kind of technology: a general purpose technology, or GPT (Goldfarb, Taska, and Teodoridis 2020; Trajtenberg 2018). GPTs are technologies that diffuse widely across a range of existing as well as new settings and uses and are typically drivers of larger transformations. One example is the steam engine, a GPT that played such a role in the Industrial

Revolution.

The term "general purpose technology" comes from the economics of innovation, where the topic has been studied in the context of such technologies' potential to produce economic growth (Bekar, Carlaw, and Lipsey 2018; Bresnahan and Trajtenberg 1995; Lipsey, Bekar, and Carlaw 1998). Although precise definitions vary somewhat, three characteristics are salient across most definitions. First, a GPT does not arrive fully formed, but rather as an improvable technology that requires further investments in research and development to reach its fruition. That investment is motivated by the second and third aspects of the definition: the technology has strong potential to be useful across a wide number of settings beyond its original purpose and also to have a range of uses within those settings. In the language of economics, these are perceived complementarities that result in potentially greater returns to investment through the recombination of aspects of the new technology with existing methods of production. In the language of the Internet, these complementarities are a result of a technology's potential to be a generative infrastructure (Zittrain 2008) whose openness allows for widespread experimentation.

These aspects of the technology lead to widespread experimentation and development of different aspects of the new technology in different settings, in part because GPTs generally arrive as clusters of technologies rather than a single artifact or process (Bekar, Carlaw, and Lipsey 2018). Moreover, the components of these clusters are often themselves general purpose technologies. This "fractal-like" aspect (Lipsey, Bekar, and Carlaw 1998) of new GPTs is integral to their adaptability, and to the recombinant innovation to which they give rise.[1]

In this chapter and the remainder of this dissertation, I describe blockchain technology as both a potential and an emergent GPT, meaning one that is early in its development relative to more established technologies, and that hasn't yet reached settlement around a single form or set of standards (Tushman and Anderson 1986; Tushman and Rosenkopf 1992). This unsettled state highlights a clear temporal tradeoff in the nature of studying technological change as it is happening: while I cede the ability to make binary arguments based on historical data about

[1] On recombinant innovation, see Stark (1996, 2009) and Lester and Piore (2004).

whether the technology is definitively a GPT, I gain the ability to use aspects of the multipart definition of GPTs that I describe above to shape my analysis of the technology as both a potential source of the transformation of practices that work across institutionalized domains as well as the creation of new alternatives.

Doing so requires first establishing the nature of the technology itself. To accomplish this, this chapter zeroes in on the clustered nature of GPTs, which I relate to blockchain technology through its roots in distributed systems engineering and cryptography (Narayanan et al. 2016). Distributed systems are generally defined in terms of 1) a set of autonomous "nodes" defined by a combination of hardware and software, each of which 2) contributes to a single "process" that 3) appears to its end users as a single system or interface (Coulouris et al. 2012; Fox 1981; van Steen and Tanenbaum 2017). Together these definitions point to a need to define the component technologies in blockchain, the processes in which they are engaged, and the means of integrating those components into a functioning system.

I approach this in the remainder of this chapter from two perspectives. The first of these is a simple descriptive accounting of the component technologies, followed by a brief description of how these are aggregated into implementations. The second is a more historical take that situates key developments in time and introduces several of the players who I will reference in subsequent chapters.

2.1 Approaching the Technology: A First Pass

While it is common to refer to "blockchain" as though it is a single technology, in practice it signifies a set of component technologies - both hardware and software - that are jointly mobilized in the creation of a distributed system that typically includes some combination of several core technologies: a peer-to-peer (P2P) network, a distributed ledger stored and managed on that network, a set of distributed protocols designed to ensure the integrity of the ledger, some form of cryptographic token, and blockchain-enabled applications based on smart contracts. After describing each briefly below, I then introduce some of the ways in which entrepreneurs recombine these components when implementing the technology.

Peer-to-Peer Network. Although such networks often fade into the background of discussions of blockchain, the necessary condition for the latter's existence as decentralized technology is a distributed set of computational nodes that jointly create and enact the technological architecture in concert with the code that runs on them.

These are many other forms of P2P network in the overall blockchain ecology. Some networks provide decentralized storage of data, while others pool their member devices' computational power and offer it for machine learning, AI and other applications. More consumer-oriented P2P networks appear as market platforms that mimic the features of existing platforms for music streaming, e-commerce, ride sharing and other applications. The "nodes" in these networks range widely from highly specialized computer hardware to consumers' mobile phones and WIFI routers.

Ledger. At its most general, the term "blockchain" refers to a ledger of time-stamped data that is replicated across some number of networked servers, so that no one entity holds centralized control over the data. This structure grew out of an older challenge in distributed systems design known as "replicated state machines," or the creation of a system of memory distributed across some number of computers (the term "state" refers to the set of known data at a point in time). Blockchains incorporate a particular form of replicated state in which time-stamped data are aggregated into "blocks," which are then aggregated into a "chain." There is some variation in the structure of these ledgers, with some storing all data on every distributed machine and others storing only so-called "shards" or subsets of the data, while other ledgers implement entirely different structures based on graph theory (e.g. the Directed Acyclical Graph or "DAG"), but they all share the fundamental condition of the absence of single, controlling entity.

The importance of such a distributed ledger is not immediately obvious until put in context. The reliability of ledgers establishing property rights has been central to the establishment of civilizations going back to Babylonia (Goetzmann 2016), though these ledgers have historically relied on the authority of the state. Ledgers have also played a fundamental role in the creation of financial and other claims, first on the state and later in domains ranging from land records to financial markets (Lamoreaux 2006; North and Weingast 1989). More recently, the rise of data as both an input and an output of digital commerce has created its own forms of establishing authoritative rights, though the solutions to date have been hampered by the infinitely copyable nature of information. The claim of blockchain ledgers to offer authoritative records, while distributing the trust that had previously been centralized in third party institutions, has thus made them an attractive alternative across a wide range of both age-old and more recent domains (Seidel and Greve 2017; Zucker 1986).

Consensus Mechanisms. Perhaps the central supporting mechanism supporting blockchains' claim to deliver decentralized and trustless interaction is their use of a combination of cryptographic protocols and other mechanisms that jointly replace negotiated or institutionalized trust with agreement by the nodes in the blockchain network. Rather than being "trustless," these consensus mechanisms in reality seek to replace trust in a third-party institution with trust in a more abstract set of coding rules (Dodd 2017; Seidel and Greve 2017).[2]

Consensus protocols are the area where the sciences of cryptography and distributed systems most clearly spill over into the applied world of actual implementations. The most commonly known - and to date, most widely applied - version of an overall consensus mechanism is Bitcoin's Proof-of-Work or PoW mechanism, which combines a much older mechanism for spam control through costly computation with a simple rule that incentivizes nodes in the network to validate the longest chain. This combination reflects Bitcoin's sole purpose of creating a decentralized ledger that enforces the scarcity and viability of the currency by controlling and immutably documenting the history of transactions, requirements that necessitate a large network with thousands of nodes.

Although Bitcoin's PoW approach is the most widely known and discussed outside of the blockchain field, the development of alternatives is perhaps the most active area of research both

[2] For an accessible discussion of this, see this Twitter thread from Cornell's Emin Gün Sirer: https://twitter.com/el33th4xor/status/1006931658338177024.

within and outside of the computer science academy. This is in part due to the extraordinary and unsustainable energy use PoW entails, but also reflects the post-Bitcoin expansion in the role of blockchain ledgers as infrastructural technology for applications well beyond cryptocurrency. Many of these applications (e.g. in banking) necessitate massive throughput of transactions rather than large-scale computation across thousands of nodes, making alternate forms of consensus protocols more relevant. These requirements also call for more significant resources in terms of processing power and connectivity than an individual, home-based user can typically provide.

These alternatives take a number of forms, though two interlinked patterns predominate. The first of these is a shift from PoW to proof-of-stake, or limiting participation in establishing consensus to network members who can bond or “stake” a significant amount of the blockchain’s token as a sign of commitment (and, potentially, risk capital that can be fined away for misbehavior). The second is a system of delegation of responsibilities to a subset of network participants who have made various commitments to the network, typically selected through some sort of voting. The combination of these two selection methods - DPoS, or Delegated Proof of Stake - is the basis for most of the more recent general-purpose blockchain systems since Ethereum (e.g. Tezos, EOS, etc.).

Smart Contracts. Smart contracts were first proposed by Nick Szabo in the 1990s in a theoretical proposal (Szabo 1997). Szabo described smart contracts as executable cryptographic code to be embedded into objects (technological and otherwise) to ensure performance of previously agreed contractual terms. The incorporation of the word “contract” has led to significant interest and critique from the legal academy. Levy (2017) points out that smart contracts omit all of the traditionally relational aspects of contracting and dispute resolution (Macaulay 1963, Macneil 1978), and also would likely have questionable standing in court, though that is far from a settled matter (see also Pasquale 2019).

In practice, such legal concerns are not always seen as the most important aspect of smart contracts; as a common saying in the field goes, they are “neither smart nor contracts.” They are instead a means of automating increasingly complex processes and memorializing their results on a ledger, all through computer code. Szabo's canonical example in his 1997 paper was not a legal setting but rather a vending machine, and explored the relationship between a consumer and the owner of the machine as a form of contracting. The setting, one in which the messier traditional aspects such as negotiation and consideration could be assumed away, is indicative of a practitioner/technologist's view of contracts that more closely reflects Oliver Williamson's notion of economic activity as "contracting" than it does a commonly understood process of contract.

The economistic sense of contracting is more closely aligned with the actual use of smart contracts in blockchain systems, given their function of automating processes that had previously occurred in other settings. One of the terms used most frequently in the blockchain ecology, and in technology systems design more broadly, is "business logic", which is typically taken to mean the rationale for and the enactment of the processes necessary to the functioning of a business (as distinct from the functioning of its technology). Smart contracts are the automating technology that most closely engages with business and other non-technological logics in the process of translating them into executable code.

Cryptographic Tokens. The dominance of Bitcoin as a frame of reference for "cryptocurrencies" has had the unfortunate effect of leading both practitioners and academics to frame the field in general in terms of Bitcoin's technology. While it is true that Bitcoin remains by far the largest cryptocurrency, in practice the past few years have see- the emergence of – depending on who is counting – eight to ten different kinds of tokens, each with its own purpose that go well beyond the original use cases of payments. The rate of change has been so significant that even the most thoughtful articles on the subject are well out of date within one or two years (Halaburda 2016).

Many of these forms involve tokenization, or creating a digital token that may circulate in markets in lieu of the object they represent. The most widely used form of tokenization operates in investment projects, which extend traditional forms of securitization into the purely digital economy (Carruthers and Kim 2011). One of the more consequential recent innovations was the introduction of the NFT, or non-fungible token structure, which is designed to be a unique, single token that represents a unique, single object or asset (thus making it tradable but technically not fungible). Originally developed in the context of what appears to be a trivial online game called "Cryptokitties," in which each animated cat had its own single token, NFTs have rapidly expanded their domain of application to include works of art, homes, derivatives, medical records and many other applications. I explore markets for NFTs and other forms of tokens in Chapter 5.

This diversity of uses and forms leads me to follow Zelizer (2011) in using the term "tokens" rather than cryptocurrency. I also define cryptocurrency as a subset of tokens in which the primary usage is for peer-to-peer payments in the mode of Bitcoin. As I show in discussing my data below, this remains a large category of projects in my data set but is less than 10% of the total, further indicating that the term "cryptocurrency" is not useful when applied to the ecology as a whole.

Composites. As mentioned above, these technologies only function as blockchain systems when they are deployed in sets that include at least two of the above. Notably, the shift of blockchain systems from being the basis for cryptocurrencies (with a unique blockchain for each cryptocurrency) to being the infrastructure for the development of decentralized applications or "Dapps" means that thousands of blockchain projects do not include a blockchain but rather build on an existing ledger. For example, "Bitcoin" is a cryptocurrency with its own blockchain, run on a P2P network of computers and using a consensus mechanism to ensure agreement across the network. By contrast, Aragon is a decentralized governance application that does not have its own blockchain but instead runs as a set of smart contracts built on the Ethereum blockchain and uses its own cryptographic token.

2.2 Approaching the Technology: A Second Pass

Having a established the components of the technology, the remainder of this chapter situates these technologies in their historical trajectories of development.

2.2.1 Precursors

The advent of the internet triggered a cultural shift in which both technologists and social theorists began to ask questions about the nature of society and the individual given the range of possibilities the internet afforded (Dyson et al. 1996; Winner 1996). Although less visible to a general audience, many of these early efforts appeared in online postings that evidenced a deep engagement with the question of how to translate foundational capitalist institutions including property rights, contracts and organizations into cyberspace. In most cases, these proposals consisted of theorizing about the ways in which recent theoretical advances in cryptography could be combined with distributed information structures and communication enabled by the internet to build parallel institutions, though in most cases the technology was not yet able to execute on these proposals.

Perhaps the most widely cited paper from this period was published in 1997 as an online post by Nicholas Szabo (Szabo 1997), introducing his concepts of smart contracts and smart property. For Szabo, smart contracts were self-executing code that could be used to control the behavior of physical objects (making the latter "smart"); in computer terms, a smart contract would embed the ability for software to effect a "state change" in the world rather than simply recording data about such a change. Szabo's canonical example - one frequently cited in subsequent papers - was a vending machine programmed with a proto-smart contract to release food only on payment of a specified amount of coins. Szabo also foresaw the rise of what we now call the "Internet of Things" by hypothesizing a second example of a leased car that could be rendered "smart" through the insertion of code into its systems that would allow the leasing company to disable the car in the event of a lack of payment. Although the paper claimed to show how software protocols could replace all stages of a contract, from meeting of the minds to performance, in reality the paper focused almost entirely on cases involving performance, and the ability to hard code rewards or sanctions depending on that performance. In so doing, it detached the notion of contract from its socio-legal moorings (Levy 2017) and instead posited something much closer to the ideal type of "contracting" from the law and economics school of thought.

Szabo wrote another post the following year that has drawn less attention, but in retrospect was remarkably prescient about many subsequent developments (Szabo 1998). The post theorized about the development of "secure property titles with owner authority," with the key problematic being how to transfer those titles "across trust boundaries" using only code rather than government institutions. Szabo defined property as encompassing identity, information and physical property, and as having a format that reflected the dominant "bundle of rights" conception of property rights as divisible. The mechanism Szabo proposed for security was based on cryptographic research published that same year, and involved a solution to a classic cryptographic trust challenge known as the Byzantine Generals Problem, though Szabo didn't specify a specific solution beyond a description of those methods. He also envisioned that property rights would be controlled by a "property club," an organization that would provide a private alternative to the state governed by a distributed process of voting and storing data on a decentralized group of servers. The paper's foresight was remarkable: the Byzantine Generals solution would later be used by Satoshi Nakamoto in Bitcoin, the property club concept foreshadowed later developments in Distributed Autonomous Organizations, and the data-based conception of property titles looked forward to today's active experimentation with state land registries.

Another early leader in theorizing about cyber-institutions was Ian Grigg, a financial cryptographer. Grigg co-led the development in the 1990s of the Ricardo trading and payment system and wrote a paper based on that experience in which he attempted to link market and state institutions as well as protocols into seven "layers" (reflecting the dominant structural metaphor for the internet) comprising a functionally integrated system of financial cryptography (Grigg 2000). Grigg introduced the idea of Ricardian contracts in that paper and later developed a fuller treatment in a paper dedicated to the concept (Grigg 2004). In contrast to Szabo's property- and performance-based conception of a contract, Grigg focused on financial instruments, which he argued were at heart promissory contracts between issuers and holders (a conception also employed by Knorr Cetina 2015). Because terms are central to those contracts, Grigg argued, an electronic translation of the contract needed to include three parts: parameters, execution code and the text of the contract itself. Although there were limited experiments with Ricardian contracts, their full implementation would have to wait for a decade.

The 1980s and 1990s were also a period of significant change in academic and industrial computer science, where the focus was generally on the underlying infrastructure and security of distributed computing. The period was particularly fertile in terms of the development of various forms of distributed technology, ranging from small networks to visions of networks so large and embedded in the built environment as to become “ubiquitous” (Weiser 1993). One of the primary concerns in this literature was how to ensure that the devices involved in this computation were able to function independently while producing trustworthy results (the consensus mechanisms I describe above are solutions to this challenge).

This consensus problem is typically framed in terms of ensuring that a system is “fault-tolerant,” with the most widely generalized fault being unpredictable or “Byzantine” behavior (Schneider 1990). This colorful term comes from one of the foundational papers in distributed systems design, which framed fault tolerance in a metaphor for the ability of a set of generals to reach agreement on the correct source of action when some unknown number of them might be traitors (Lamport, Shostak, and Pease 1982). This “Byzantine Generals Problem” was proven by the authors to be solvable, under a specific set of conditions, only when less than one third of the generals were “Byzantine,” which is to say when less than one third of the devices in a system are unreliable.[3] Systems robust to such errors are said to be “Byzantine Fault Tolerant.” In keeping with the probabilistic nature of the proofs of these systems’ robustness, as well as rapid change in computational power, the development of fault-tolerant consensus mechanisms is one of the most active areas of research in computer science.

[3]This solution is predicated on the impossibility of two generals with opposing views to reach agreement if they couldn’t trust the view of a third, *inter alia*.

2.2.2 Bitcoin and Its Blockchain

It is customary to refer to the pseudonymous Satoshi Nakamoto as the ur-figure of cryptocurrency and blockchain. In a 2008 listserv email that was subsequently formalized into a paper (Nakamoto 2008), Nakamoto proposed a new system that used a combination of game theoretic incentives, distributed peer-to-peer networked software and cryptography to create a ”peer-to-peer electronic cash system.” The result was Bitcoin, the first widely successful cryptocurrency, with a total market value of just over $ 1 trillion as of October 8, 2021.[4]

In practice, Nakamoto engineered an innovation that drew on multiple environmental building blocks. Narayanan and Clark (2017) describe a broad base of prior academic research in multiple domains from which Nakamoto worked, key among which was cryptography. As an applied field of mathematics, cryptography is based on mathematical proofs, and emphasizes the protection of data and communications from attackers (Katz and Lindell 2015). Nakamoto also drew on distributed systems engineering by designing Bitcoin's code to run on a distributed network of computers and servers. Bitcoin's use of the computationally intensive "proof-of-work" method of consensus relying on work done by a dispersed network of computers to secure the data on the Bitcoin blockchain (known as ”mining”) reflected this grounding in both cryptography and distributed systems engineering.

Nakamoto's practices also reflected a commitment to open source development that laid the groundwork for those who followed. The Bitcoin source code has been available since 2009 on Github, which has also served as the primary platform for its communication and management of its development.[5] The use of Github has since become among the most powerful norms in the blockchain ecology.

The open source community that ultimately grew around Bitcoin differed, however, from the more commons-oriented peer production (Benkler 2002, 2006) that initially marked other open source efforts. Nakamoto's design created direct and indirect incentives for all holders of Bitcoin,

[4]https://web.archive.org/web/20211008141230/https://coinmarketcap.com/currencies/bitcoin/

[5]The core Bitcoin repository is located at https://github.com/bitcoin/bitcoin. Early versions of the source code are available at http://satoshi.nakamotoinstitute.org/code/

which included economic incentives for providers of computing power to secure the network, as well as more diffuse incentives for engineers writing code who hold Bitcoin and therefore want to see its software maintained. In spite of its stated antipathy to authority and hierarchy, the broader community of Bitcoin developers has also followed the broader open source community's implicit solution to the problem of quality control (Benkler 2002, MacCormack et al 2012). Although technically anyone can contribute to the codebase, those contributions remain suggestions until they are discussed, vetted and incorporated by a much smaller group of core developers. This set of ”core devs” form a centralized set of decision-makers who govern day-to-day decisions about the codebase.

The success of Bitcoin opened the door to an early wave of experimentation and hundreds of new cryptocurrencies that followed its precedent in the following years. These are often referred to in the community as the era of "Blockchain 1.0", referring to the initial configuration of a cryptocurrency linked to its own blockchain used only to secure the cryptocurrency. This experimentation was ultimately limited, however, to cryptocurrencies themselves rather than the underlying blockchain. This limitation was due in part to the code's pure focus on transactions and in part due to the changing composition of the broader Bitcoin community as it grew to include more non-technical members. These disagreements led to the departure from that community of many technologists who would go on to be instrumental in the founding of a number of new blockchain projects.[6]

2.2.3 Ethereum, Smart Contracts and the DApp Economy

Among the most consequential of those technologists was Vitalik Buterin, who developed the underlying protocol for Ethereum by copying and modifying Bitcoin's protocol in several ways. Buterin and the project's other founders envisioned the Ethereum blockchain acting as an enabling, generative platform for third parties to develop new applications (Buterin 2013; Wood

[6]Mike Hearn, the co-creator of R3's Corda platform for the finance industry, has been among the most vocal of these technologists. See https://blog.plan99.net/the-resolution-of-the-bitcoin-experiment-dabb30201f7#.h81ihjioy for more detail.

2014). To accomplish this, the founders envisioned a transformation in the purpose of a blockchain ledger – rather than securing a single cryptocurrency, as in Bitcoin, Ethereum's P2P network would aggregate its computational power to execute the encoded agreements underlying all of the applications based on Ethereum's blockchain. Just as individuals were free to enter the P2P network to run nodes, developers of decentralized applications (or "dApps") would also be free to develop applications to run on Ethereum's technological platform, making it doubly permissionless.

Rather than being a means of generalized peer-to-peer payment like Bitcoin, the Ether token was designed to be the medium used by applications to pay for their use of the Ethereum computational infrastructure. Buterin also moved away from Bitcoin's precedent by funding the platform's development through an initial coin offering (”ICO”) in which units of the Ether currency were sold to the public.[7] In keeping with standard practice in the open source community (O'Mahony 2003; O'Mahony and Bechky 2008), proceeds were used to fund the creation of the Ethereum Foundation, which employs many of the Ethereum ”core devs” and governs both the ongoing development of Ethereum's core code as well as its relations with the broader ecosystem of projects that has grown from the Ethereum blockchain.[8]

Key among their innovations was the development of a new general-purpose coding language written to enable *smart contracts*, or "agreement[s] whose execution is both automatable and enforceable" (Clack, Bakshi, and Braine 2016, p.2). Ethereum's emphasis on smart contracts, and in particular on standardizing smart contracts that allow for the creation of ICOs for non-Ethereum projects on the Ethereum blockchain, provided the basis for much of the subsequent explosive growth in blockchain projects. Many projects planning to develop their own blockchain platforms have capitalized on Ethereum's standardization of issuing new tokens by raising funds first on Ethereum and then using those funds to build their own infrastructure.

The rise of this platform structure has given rise to a second category of infrastructural

[7] Ethereum's ICO was precedent-setting in many ways, but was not the first. See https://web.archive.org/web/20210618113547/https://www.forbes.com/sites/laurashin/2017/09/21/heres-the-man-who-created-icos-and-this-is-the-new-token-hes-backing/

[8] https://www.ethereum.org/foundation

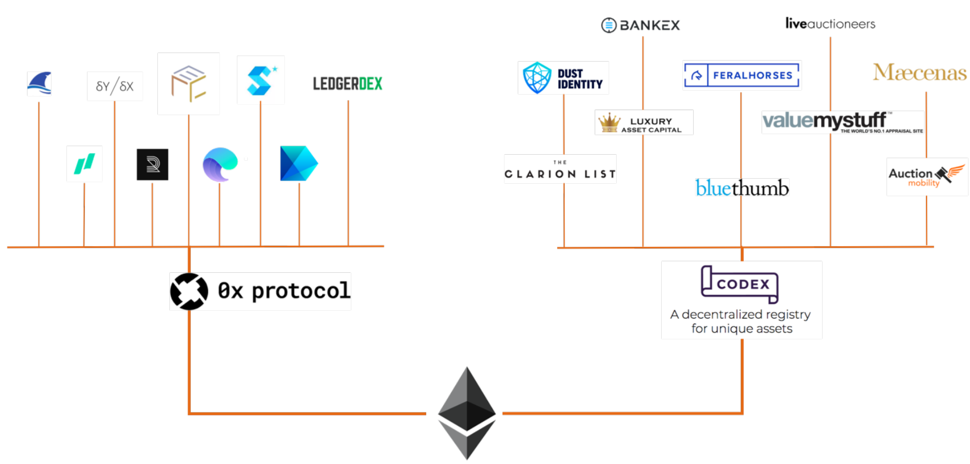
projects. Built to draw on the computational and recording functions of infrastructural blockchains such as Ethereum, these protocols form the basis of platforms for the further development of applications and implementations in a particular functional domain. As a result, they are essential to the expansion of blockchain technology's span of influence beyond cryptocurrencies, and thus to its nature as a general purpose technology. One prominent example is the 0x protocol for the decentralized trading of cryptographic tokens, which draws on the settlement finality and other functionality of the Ethereum blockchain. The protocol - named for the first two alphanumeric digits in a blockchain address - was used by 45 other projects as of February 2020[9].

Figure 1 provides a visual example of the kind of protocol-based layering I describe above. The base of the graphic uses the Ethereum logo to represent the use of the Ethereum blockchain (and thus its blockchain-layer protocol) by the constellations of projects whose logos appear above it. These projects in turn are built on two intermediary or blockchain-enabled protocols, the first of which is the 0x protocol described above. The second is the Codex Protocol, developed to support the art world by providing blockchain-based records of provenance secured by the Ethereum blockchain. These two blockchain-enabled protocols each build domain-specific (but still open) functionality onto the general purpose Ethereum blockchain, and are used for this purpose by dozens of other domain-specific applications and projects whose logos appear at the top of the image.

Since Ethereum's launch, this layered, platform conception of blockchains has become a dominant one in the domain of public or permissionless blockchains, and has structured the evolution of subsequent projects. Some of these projects have built alternatives to Ethereum with the same core objective of providing application infrastructure but with different technological and governance parameters. Others are building new technologies that boost the transactional capabilities of existing blockchains (primarily Bitcoin and Ethereum), either by reducing the data demands on the underlying chains (defined as "scaling") or by enabling computation to occur

[9]https://github.com/0xProject/wiki/blob/master/community/2-List-of-Projects-Using-0x-Protocol.md

Figure 2.1: Layered Blockchain Projects and Protocols



across multiple chains ("interoperability"). Moreover, each of these technologies often has multiple projects building applications that use the new technology as an infrastructural layer for their own development.

2.2.4 Enterprise Platforms: R3 and The Linux's Foundation's Hyperledger Project

R3 was launched in the fall of 2015 as a consortium of banks, stock exchanges and other organizations from the financial community, and by late 2016 included effectively every major bank in the world. The organization R3 CEV was formed later that fall, with a technology team led by Richard Gendal Brown, a former senior technologist from IBM, as well as Mike Hearn (formerly a Bitcoin Core developer) and Ian Griggs (the inventor of Ricardian contracts). Although R3 suffered some high-profile defections in late 2016, when Goldman Sachs and two other banks departed the consortium, it remains by far the largest single-industry group working on DLT.

R3's core offering is the Corda system, a "blockchain-inspired system" created expressly for the financial industry. Corda was designed to enable consensus within transactions rather globally, to keep information private, and to link performance features of smart contracts to the legal protections and dispute mechanisms of Ricardian contracts (Brown et al. 2016). These features accord with R3's constituent base of highly regulated institutions, which required a combination of privacy and customization unavailable on public chains. The Corda software is open source and has been designed to support the development of trading and other protocols by its potential users. It also includes "notaries," which perform a similar function in Corda as in Ripple's Interledger Protocol – that of a trusted third party whose intercession replaces computationally intensive distributed proof-of-work consensus mechanisms.

R3 was also a founding member of the consortium Hyperledger Project, which was announced by the Linux Foundation in late 2015 as an umbrella meta-project whose members are largely established firms with goal of shepherding and governing the open source development of blockchain protocols. Hyperledger was launched in 2016 as an incubator for various projects using combined code from IBM, Intel, Microsoft, Digital Asset Holdings, SWIFT, JP Morgan and other firms. Hyperledger's various platforms are designed as open source codebases that corporate and other users can adapt to create their own networks and blockchains. Unlike Ethereum, there will be no single public Hyperledger, nor is there a Hyperledger currency. Instead, Hyperledger as a project was conceived to develop private or permissioned ledgers for corporate users prioritizing confidentiality and local regulatory compliance over full decentralization.

2.2.5 Parallels and Convergences

At the time of this writing in 2021, two broad patterns are generally characteristic of the current trajectory of implementations of the technology. I term the first *parallel development*, by which I mean the simultaneous development of implementations by established firms and government institutions and by challenger startups. This pattern is not immediately evident when looking at implementations from the perspective of "ICOs" (which definitionally captures only startups) or the enterprise consortia I describe in the preceding section, but it becomes clear when looking across my entire set of projects.

The pattern is clearest in the domain most widely associated with the technology, which is to say fintech. Wall Street banks and regulators were among the first incumbent institutions to experiment with the technology (primarily for record-keeping at first, and more recently for bond issuance), and remain among the most active. At the same time, the Bitcoin model of cryptocurrency as a payment medium has since been succeeded by waves of innovation built around the issuance, trading, portfolio management and custody (or storage) of digital assets.

While these developments are ongoing, perhaps the most surprising development in late 2020 and into 2021 has been the simultaneous attempt to develop digital money designed to maintain the price stability (and thus usefulness for payments) of traditional currency. Rather than Wall Street, however, the primary incumbent actors in developing these payment media are the world's central banks, dozens of which are engaged with pilot implementations of digital versions of their national currencies (Auer, Cornelli, and Frost 2020). These Central Bank Digital Currencies (or CBDCs) are in various stages of development but are generally being developed with an eye to retaining a role for central bank policy in an increasingly digital financial system (Auer and Böhme 2020).

The primary threat cited by most of these banks is not traditional cryptocurrencies such as Bitcoin, but rather so-called stablecoins. These are a subset of cryptocurrencies that are also used for payments but are designed to maintain a stable monetary value (typically as a peg to a traditional currency or gold) while also retaining the distributed, P2P character of cryptocurrencies. According to blockchain information media company The Block, the aggregate market value of issued stablecoins grew from $5.9 billion at the start of 2020 to more than $106 billion in mid-June 2021.[10] Facebook's highly publicized stablecoin Diem (formerly Libra) and its potential implications for the financial system brought the attention of regulators and legislators to stablecoins, which at the time of this writing has produced at least one proposed piece of legislation to limit their potential impact.[11]

[10] Data archived as of June 16, 2021: https://web.archive.org/web/20210616164120/https://www.theblockcrypto.com/data/decentralized-finance/stablecoins/total-stablecoin-supply-daily

[11] See the Stablecoin Tethering and Bank Licensing Enforcement (STABLE) Act at https://tlaib.house.gov/sites/tlaib.house.gov/files/STABLEAct.pdf.

The other broad pattern evident in my data is the *convergence* of blockchain and other potential digital transformation GPTs. Like parallel development, convergence operates across both incumbents and startups, though it differs in that it is more limited in its span of domains. The primary complements to blockchain in this pattern are the Internet of Things (IoT) and artificial intelligence/machine learning (AI/ML). The former is generalizable as the use of distributed sensors embedded in physical systems such as supply chains or energy grids. These sensors collect data that, when aggregated and stored on a blockchain, allows for the tracking of goods across their lifecycle. AI and ML, by contrast, are technologies of large-scale computation, and feature most prominently in blockchain-enabled markets for data, and in applications for bots in marketplaces.

Chapter 3

Identifying an Emergent General Purpose Technology: An Interpretive Data Science Approach to Classification

I argued in the prior chapter that approaching blockchain technology as a GPT provides a potential structure to the larger question of digital transformation. The very generality of the technology poses analytical challenges: if a technology can be and become so many different things, how is it possible to study it as a whole? This chapter provides a methodological grounding for the chapters that follow by developing a classification system based on actual implementations of blockchain technology as opposed to its potential uses. Doing so allows me to establish a macro-level view of the breadth and mechanisms of the technology's use that I explore in greater detail in each of the subsequent chapters. I develop this framework by beginning with insights from both the economics of innovation and the sociology of classification and categories, which I marry with large-scale text analysis to develop my own interpretive data science methodology.

3.1 Identifying General Purpose Technologies: The Dual Challenge

The multi-part definition of general purpose technologies I described in the prior chapter implies two closely related challenges to identifying a GPT in the present. The first part of this dual challenge is deceptively simple, namely the identification of the GPT itself. The articulation of the technology's components I develop in Chapter 2 is a necessary step toward this articulation but is not sufficient in itself. Establishing a GPT requires identifying *both* the components and their generality of purpose.

Much of the work on GPTs in economics and elsewhere has focused on the second aspect of the challenge, namely the identification of the breadth of uses of the technology. Scholars in this vein have generally relied on historical data and the existence of authoritative taxonomies against which to measure the historical diffusion and effects of a given study's focal technology. One standard empirical approach in this literature is to use patent filings to answer whether a given focal technology is a GPT, as well as to measure the impact of that technology more broadly (e.g. Feldman and Yoon 2012; Hall and Trajtenberg 2004). At the most general level, they do this by embedding the technology being studied into one of two closely related taxonomies. The first is the system of patents and their classifications. The conditions for granting a new patent include that the technology is an advance, and that it references - but is different from - all "prior art" or previous advancements on which it builds. A closely related part of the process is the classification by patent office experts of the new technology into an established taxonomy of technologies. In the U.S. and Europe, this taxonomy is the Cooperative Patent Classification (CPC) system. Studies of GPTs using patent data have generally used some combination of the citation networks between individual patents as well as the timing and taxonomic location of those patents to measure the diffusion and temporal effects of the technologies they study.

Another taxonomy used by many studies is the closely related taxonomy of the North American Industry Classification System (NAICS), a taxonomy of organizations categorized by their primary form of production. NAICS terms these “establishments,” which it distinguishes from “enterprises” by the fact that enterprises can include multiple establishments, while establishments are constrained to a single classification. Studies of GPTs using the NAICS taxonomy generally begin with technology patents, and then take the organizations to whom the patents are assigned as their unit of analysis, along with NAICS classifications of those organizations. This allows researchers to measure economic impacts using those and related companies’ financial performance.

Table 3.1: Established Taxonomies for Technology and Industry Classification

CPC Technology Taxonomy		**NAICS Production Taxonomy**	
code	Technology Class	*code*	Industry Sector
A	Human Necessities	11	Agriculture, Forestry, Fishing, Hunting
B	Performing Operations; Transporting	21	Mining, Quarrying, Oil & Gas Extraction
C	Chemistry; Metallurgy	22	Utilities
D	Textiles; Paper	23	Construction
E	Fixed Constructions	31-33	Manufacturing
F	Mechanical Engineering; Lighting; Heating; Weapons; Blasting	42	Wholesale Trade
G	Physics	44-45	Retail Trade
H	Electricity	48-49	Transportation and Warehousing
Y	Various Technology	51	Information
		52	Finance and Insurance
		53	Real Estate, Rental, Leasing
		54	Professional, Scientific, and Technical Services
		55	Management of Companies and Enterprises
		56	Admin., Support, Waste Mgmt. & Remediation
		61	Educational Services
		62	Health Care and Social Assistance
		71	Arts, Entertainment, and Recreation
		72	Accommodation and Food Services
		81	Other Services (ex Public Admin.)
		92	Public Administration

The prevalence of these taxonomies as conceptual infrastructure for this literature would seem to imply that they also provide a framework for the study of digital transformation. Unfortunately, their usefulness for historical studies does not extend to the study of novel technologies. While systems of classification are a necessary condition for identifying these technologies, taxonomies such as NAICS are well-suited to the circumstances out of which they evolved, but evolve slowly over time and can only include new technologies within their existing framework of categories. This makes them only partially suited to the dual challenge I describe above given that blockchain's potential lies in a combination of existing categories as well as new ones created by the technology.

These shortcomings point to the need for an approach to classification grounded in the uses of the technology itself, and in their hierarchical taxonomic relations, rather than simple extrapolation from existing frameworks. The operative questions then becomes how to identify and classify these uses and relations. I turn to recent developments in the sociology of categorization to identify the components of the process I develop.

3.2 Classification and Taxonomies

Sociological research on the importance of categories in structuring social activity is rooted in the core of the discipline, in Durkheim's foundational analysis of the articulation of new and distinct roles inherent in the Industrial Revolution (Durkheim 1997). In its more contemporary form, foundational work on categories in economic sociology identified them as constraints that disciplined market actors by providing frameworks against which audiences judge legitimacy (Zuckerman 1999). The role of categories in this work necessarily rests on their reification in existing taxonomies that provide legitimation to social actors, and scaffold analysis by providing the implicit levels that the researcher holds constant while focusing on those actors' behavior at a single level (e.g. Zuckerman's analysis of stocks in the context of their industry codes in the Standard Industrial Classification (SIC) codes, the precursor to the NAICS). Even studies that trace the development of new categories typically do so within previously defined markets (Khaire and Wadhwani 2010) or extensions of them (Navis and Glynn 2010).

More recent work on categorization has theorized a potential escape from this constraining role of categories and, by extension, of the empirical constraint of reliance on existing categories. This work focuses on the agency of entrepreneurs (broadly defined) in crafting narratives about their innovations (Kennedy, Lo, and Lounsbury 2010; Lounsbury and Glynn 2001; Ruef 2000; Wry, Lounsbury, and Glynn 2011). While these narratives draw on existing frameworks for legitimacy, they also provide a means for entrepreneurs and technology developers to associate their innovations with aspirational expectations and imagined affordances (Durand and Paolella 2013; Glaser, Krikorian Atkinson, and Fiss 2019) that look beyond existing categories. Environments in which both the entrepreneur's project and the market it is involved in creating are new, the "categorical imperative" identified by Zuckerman's early work shifts from one of alignment with institutionalized categories to one in which the entrepreneur and the audience are engaged in a mutual process of sensemaking (Weick 1995) and sensegiving about emergent categories (Durand and Khaire 2017, Kennedy 2008, Navis and Glynn 2010).

This work points to organizational texts as a rich source of data for the study of emergent categories, though it doesn't address the central question of how to build a new classification system. Building such a system, argues Starr (1992), requires four interlocking decisions about 1) the core principles and domain of classification, 2) the assignment of items to categories, 3) the naming of categories and 4) the ordering of these categories into hierarchies.

I establish the domain of classifications – implementations of blockchain technology – in earlier chapters. For the core principles, I turn to recent work on classification by Hannan et al. (2019), who separate categories from the more abstract concepts they embody. Concepts in their framework are abstract representations composed of sets of features or characteristics of some kind of object. These sets appear as coherent structure within the larger blank canvas of all possible combinations of features. Given this approach, it is not surprising that the authors point to topic modeling as a particularly useful method to use in classification.

Topic modeling is a large-scale machine learning approach for working with large amounts of text (Blei, Ng, and Jordan 2003; Hannigan et al. 2019). Similarly to Hannan et al.'s definition of concepts as abstractions, topic modeling is grounded in the assumption that conceptual structure is latent and abstract, and that it can be identified through the probabilistic analysis of how words appear together in high-dimensional spaces. The core concept of topic modeling is that the words being used define the semantic space of all possible combinations of those words, and that the topics that cohere (as sets of words) within that space are probabilistically more likely than other sets.

The second set of decisions involves the assignment of items to categories once the classification system is defined. Bringing this process onstage requires a means of identifying a viable link between topic modeling output and the documents the models index. I make this link using another output of topic modeling, which is the probabilistic measure of how much of a given document's content is captured by a given topic (or, in my parlance, concept). A higher value on this metric for a topic-document pair indicates a higher model estimation of the prevalence of that topic or concept's set of features (or words) within a given document. To preview the method I develop below, I use these numbers as the basis for assigning documents to topic-concepts; groups of documents thus assigned then become observable potential categories.

Starr's (1992) identification of naming as a necessary step foregrounds one of the most profoundly interpretive aspects of topic modeling, namely the need to *name* topics that appear only as probabilistic lists of words. Although it is possible to use new algorithms for this (see e.g.), many analyses leave the naming to the researcher based on their close interpretation of the topical content. This latter approach is closely aligned with the longer tradition of interpretive coding in qualitative research, which provides an established set of practices on which I draw (Glaser and Strauss 1967; **grodal_achieving_2020**; Tavory and Timmermans 2014).

There remains one challenge to be solved in developing a taxonomy, namely how to identify both the levels of the taxonomy and the vertical linkages across levels that make it a hierarchy. The methodological challenge here is creating what Barley (1990, p. 66) calls "an explicit conceptual lattice for moving between adjacent levels of social organization."

Given my use of topic modeling, the question becomes whether any one topic model will be capable of identifying multiple *levels* of concepts within the same analysis. While hierarchical topic models do exist, they generally require each document to only a single component of the hierarchy. This constraint is excessively restrictive for an analysis of an emergent phenomenon, and is particularly ill-suited for my data (as I describe further below). An alternative approach would be to search for clustering within topics or with the vectorized corpus vocabulary, though this approach would necessitate still further interpretive choices about the clustering algorithm and the ex ante selection of the number of clusters.

Instead, I start from Mohr and Duquenne's insight (1997) that the same data set can provide material at multiple levels of analysis. I wed this insight a logical observation that a topic model constrained to a relatively small number of topics will identify topics that represent fewer, broader concepts. Conversely, topics should become increasingly specialized and narrow as the number of topics in a model grows. By extension, two models using different numbers of topics *on the same corpus of documents* should identify similar topics but with varying levels of granularity, with the "larger" model (with more topics) identifying both more and more detailed information. I take this approach by constructing a single corpus and vocabulary, then using multiple topic models varying in their numbers of topics to identify the nested levels of the emergent conceptual hierarchy of classifications within my corpus of documents.

Although the existing taxonomies have up to six levels (and my data could potentially support that level of detail), I judge three levels of nesting categories to be sufficient to capture the diversity in both the range across application sectors as well as the depth of variety within them.[1]

3.3 Data and Methods

I take an interpretive data science approach (Hannigan et al. 2019; Nelson 2017) to identifying underlying concepts through the semantic space of documents produced by projects themselves, as well as the relationships between those objects and concepts. I approach this abductively by developing a process that does not separate interpretation from the analytical work, but rather weaves both together as ongoing elements of an iterative process (Tavory and Timmermans 2014). This interpretive data science requires a deeper level of engagement with

[1] Elinor Ostrom (2005) also recommends three levels as the appropriate amount for analysis in institutional settings.

both the underlying texts, their individual content, and finally, with the outputs of models in order to render the kinds of insights needed here. My particular approach to this combination draws on the algorithmic output of topic models as well as a qualitative approach to naming topics based on established qualitative methods for thematic coding and categorization (**grodal_achieving_2020**).

I follow the broad methodological outline developed by Hannigan et al. (2019) by developing the data and methods across three stages, though what follows below is necessarily a condensation of a more complicated process. I present that process in greater detail in Appendix A.

3.3.1 Data Collection and Rendering

I base my analysis on a data set I have constructed of roughly 5,000 projects implementing aspect(s) of blockchain technology that I identified these through social media posts, ICO websites, industry media, public announcements and other sources. From these projects, I construct a unique data set of roughly 5,500 project documents. These documents are as diverse in format and intent as the projects they describe. They include documentation of pilot projects, open source protocols and other efforts by global corporations, central banks and other actors in addition to the documents of startups that participated in the ICO boom. The norm of crafting a white paper to introduce new projects was also established by Bitcoin's pseudonymous founder(s) Satoshi Nakamoto (Nakamoto 2008), and is sufficiently established that nearly every project publishes at least one such paper that describes some combination of the project's technology, business objectives, mission, community strategy and other factors. These papers reflect the melding of social and technological organization in blockchain projects in that they blend the technological practice of producing technical papers introducing new technology with the entrepreneurial practice of writing and circulating business plans (Baden-Fuller and Morgan 2010; Ghaziani and Ventresca 2005). In contrast to privately circulated business plans, however, blockchain project documents are publicly posted online. They are a central point of references for the various external stakeholder communities with which projects interact.

As part of the process of documenting these projects as I gathered documents, I also generated a preliminary categorization of each project based on manual coding. While not a primary input into the present analysis, this approach provided two benefits. First, it required me to engage with each project in the context of the larger ecology and to grapple with the breadth of possible categorizations. More prosaically, it allowed me to identify and remove a small set of 23 projects that were so specialized that they lacked similarity with any other project (and thus couldn't be classified as part of a taxonomic group).

Rendering such heterogeneous documents into data in a format amenable to analysis is the next challenge. Reading the texts into R appears deceptively easy but in practice requires significant extra pre-processing beyond the methods traditionally used in text analysis. In some cases, this is due to the formatting of documents into multiple columns, while in others it is the result of the documents being created by actors working in different technical, industrial, national and ideological contexts. The relative novelty of the technology also makes existing dictionaries for lemmatization less useful than they might otherwise be. I address this by creating a custom set of 11 dictionaries that I apply to the papers using the *stringi* (Gagolewski et al. 2018) and *quanteda* (Benoit et al. 2018) packages in R. I describe these libraries in Appendix A.

3.3.2 Running Topic Models and Rendering Output

Given the need to work across multiple models, at multiple levels, my approach to topic modeling differs from the standard approach. I run a broad range of topic models differing solely in the number of topics they estimate and bound this range through a combination of the precedent set by existing taxonomies (at the bottom end of the range) and my preliminary findings from manual coding (at the top end), as well as the general approach in the computer science literature of using a multiple of the number of "true" topics in estimation (Airoldi et al. 2010; Lu, Mei, and Zhai 2011; Xie and Xing 2013). I thus run a series of models with the number of topics ranging from 25 to 1,300.

I generate the individual models within this range using correlated topic models (Blei and Lafferty 2007), or CTMs. Unlike the Latent Dirichlet Allocation approach used by most analyses using topic models, CTMs do not assume that each topic is independent of the others. Instead, they use a statistical structure that allows for correlations between topics. This makes them both more relevant for my data, given the likelihood of clustering on multiple dimensions, as well as for my analytical question. I use the stm package in R without covariates, and have used Spectral initialization to keep my analysis replicable (Roberts et al. 2018). Spectral initialization also facilitates the comparisons across models.

3.3.3 Creating Taxonomy

I construct the taxonomy using the labeled topic models, beginning with the selection of models at each of the three levels. In order to avoid over-reliance on any single model, I choose three for each of the three levels of hierarchy in the taxonomy, for a total of nine models. I anchor the most general, highest level of the hierarchy in the three smallest (in terms of the number of topics) models, and use the three largest as the basis for the bottom and most detailed level of the taxonomy. As a check on over-reliance on my own interpretation, I use three models for the middle level of the hierarchy whose numbers of topics were chosen by algorithm. As an additional check, I search for agreement across the models within a given level in order to identify formal categories. Specifically, I only move forward with topics present in at least two of the models at a given level, and term these *concordance* topics.

Given these concordance topics at each of the three levels, the penultimate stage of the analysis is identifying the linkages between the levels that constitute the hierarchy. I draw on the established approach of multimodal or affiliate networks in social network analysis. Like the affiliate networks used elsewhere (e.g. Mohr and Duquenne 1997), I use *co-occurrences*, or in my data's case, the joint assignment of a given project to topics at multiple levels as evidence of a link between the categories involved. For example, a project assigned to a topic about trading in a more generalized model, as well as a topic about exchange infrastructure in a more granular model (one with more topics), would indicate a hierarchical linkage between trading and exchange infrastructure as categories.

The final step in this stage is making the transition from named and assigned topics to actual categories. While the gap between the two was most often minimal, some cases necessitated decisions about what to incorporate as a category and what not to. My approach to such cases draws on the availability of comparative (across models within a level) and relational (across levels) context for each topic. This stage of the process drew heavily on both the computational outputs of the topic models as well as interpretation that enabled this comparative and relational analysis, and also included several stages of disambiguation to establish categories in cases where the categories appeared inconsistently across topics.

3.3.4 Computational Disambiguation

Perhaps the most important step in disambiguation involved the two best-known categories in my data set – blockchain itself, as well as cryptocurrency. These were the two least consistent in their manifestations across models at the Mid Level. One of the Mid-Level models included a generic topic about blockchain, in addition to a clear *Blockchain as infrastructure* topic, but had no obvious *cryptocurrency* topic. This seemed to reflect the fact that the Top-Level **Cryptocurrency** topic also incorporated nearly as many terms about cryptocurrency blockchains as it did about currency itself, but wasn't a pattern that repeated in the two other Mid-Level models. In those, meanwhile, *Cryptocurrency* was instead almost indistinguishable from *Crypto payments*.

This ambiguity did not appear in the Top-Level topics relating to cryptocurrency and blockchains, where concordances were clear and easy to identify. That said, the content of these topics often overlapped at enough points to make the patterns at the Mid-Level particularly challenging. This is largely because the technology – even in the most general, Top-Level topics – had several aspects around which categories coalesced. The first of these was the computational work of producing and securing individual blocks of data, typically offered as a pooled third-party service of mining or block validation for various blockchains; these are gathered in the Top-Level **Blockchain Production** topic. The Top-Level **Cryptocurrency** topic blended terms describing both cryptocurrency and the blockchains on which they are produced, secured and transacted. The blockchain technology language in these topics is also consistent in its emphasis on the Proof-of-Work methods used by Bitcoin, which are also dominant (or were at the time of my data gathering) among cryptocurrency-specific blockchains. These are distinct from the infrastructural aspect of general-use blockchains discussed in the Top-Level **Blockchain as Infrastructure** category in that each cryptocurrency blockchain is specific to an individual cryptocurrency, while the latter are used for other purposes.

There are thus two questions to disambiguate here. First, what role does blockchain technology play in the Top-Level topics given that Production and Cryptocurrency overlap? This could be answered by disambiguating the lack of concordance in related Mid-Level topics, as could the distinction between cryptocurrency and crypto payments. Table C.1 below is a first step in that process, and lists representative terms from each of the topics, for each of the Mid-Level models, that required further analysis.

Table 3.2: Content of Ambiguous Mid-Level Blockchain and Cryptocurrency Topics

Mid-Level Model	**Topic**	**Representative Terms**
1	27	protocol; layer; offchain; decentralized; centralized; onchain; neo; domain; usecase; trustless; built; opensource; native; discovery; metadata
2	39	dapp; layer; developer; application; module; api; neo; blockchain; template; framework; sdk; architecture; blockchainnetwork; eos; interface
3	18	smartcontract; decentralized; blockchaintechnology; ethereumblockchain; aim; opensource; adoption; neo; legacy; vision; technological; economy; ethereum; blockchain; centralized
1	69	api; web; developer; browser; software; backend; application; javascript; microservice; plugin; opensource; frontend; sdk; app; integration
3	62	module; api; interface; template; integration; backend; component; developer; application; microservice; tool; workflow; framework; custom; plugin
1	95	dapp; mainnet; eos; sidechain; eosio; bp; testnet; ram; developer; elastic; virtualized; eostoken; launch; deploy; inflation
3	15	dapp; eos; blockproducer; mainnet; dpos; delegate; eosio; interchain; sidechain; bridge; native; chain; bp; governance; crosschain
2	10	contract; oracle; derivative ; relayer; onchain; rebalance; offchain; proxy; smartcontract; call; agreement; logic; execute; smart; contractual
3	14	contract; oracle; smartcontract; offchain; onchain; smart; relayer; derivative; logic; execute; agreement; chainlink; call; template; ricardian
1	80	contract; smartcontract; ether; gas; ethereum; ethereumnetwork; smart; ethereumblockchain; solidity; executed; execute; evm; code; erc20token; deploy
2	75	ether; ethereum; ethereumblockchain; ethereumnetwork; erc20token; erc20; eth; smartcontract; gas; ethereumbased; metamask; solidity; crowdsale; deploy; ethereumplatform
3	77	ether; bounty; bug; ethereumnetwork; ethereum; alpha; hunter; ethereumblockchain; inc; github; metamask; erc20token; beta; testing: release
1	37	miner; proofofwork; pow; difficulty; bitcoin; mine; nakamoto; bitcoinnetwork; puzzle; attacker; hashing; mining; blocksize; block; asic
2	82	miner; pow; proofofwork; difficulty; block; blocksize; attacker; newblock; hashing; bitcoin; bitcoinnetwork; nakamoto; mine; puzzle; previousblock
3	3	miner; bitcoin; proofofwork; mine; hashpower; pow; mining; bitcoinnetwork; difficulty; nakamoto; asic; hashrate; satoshi; puzzle; hashingreward
1	4	block; newblock; signer; previousblock; validator; blockheader; header ; tree; batch; height; hash; invalid; blockhash; validation; genesisblock
3	55	block; newblock; attacker; tip; previousblock; dag; nxt; forge; blockgeneration; pow; interval; attack; nextblock; tangle; weight
1	84	blockchaintechnology; blockchain; centralized ; blockchainbased; peertopeer; decentralized; adoption; secure; technological; technology; blockchainnetwork; innovation; decentralization; intermediary; revolution
2	58	cryptocurrency; digitalcurrency; bitcoin; blockchaintechnology; currency; satoshi; peertopeer; litecoin; altcoin; revolution; money; world; paymentsystem; nakamoto; ripple
3	70	payment; transaction; transfer; digitalcurrency; currency; transactionfee; stellar; wallet; money; gateway; transact; paymentsystem; traditional; peertopeer; instant
1	23	card; remittance; debitcard; creditcard; atm; banking; fiat; merchant; visa; paymentsystem; payment; crypto; bank; fiatcurrency; cards
2	83	crypto; card; fiat; debitcard; wallet; atm; cryptocurrency; fiatcurrency; cryptocurrencyexchange; cold; visa; multicurrency; withdrawal; instantly; instant
2	32	remittance; banking; bank; unbanked; centralbank; africa; african; financialservice; cash; philippines; financialinstitution; money; gdp; cryptobank; inclusion
3	81	crypto; cryptocurrency; debitcard; atm; card; fiat; cryptocurrencyexchange; wallet; cryptocurrencymarket; altcoin; fiatcurrency; banking; cold; visa; referral

As the table shows, there is significant overlap between the content of most of these topics, which makes assignment and classification based purely on interpretation more subjective than in most other areas. Even general-purpose topics at the Mid-Level tend to align more cleanly across the three Mid-Level models (examples available on request). While it would seem to make sense to look at the relationships between assignments for topics, that method is not available here given that each model represents a different set of topics, assignments and projects.

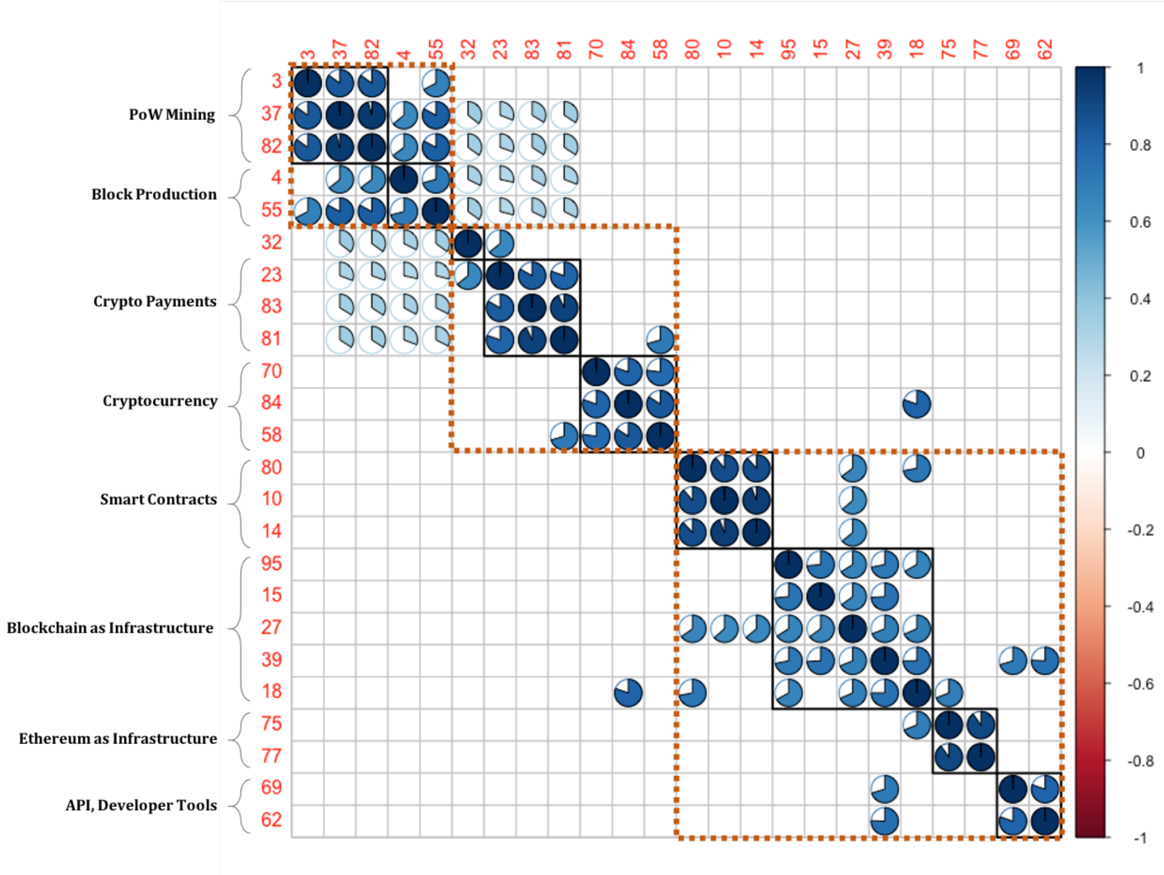
I address this by exploiting the fact that the models all used the same vocabulary across all models, regardless of differences in the numbers of topics and projects that remained after the cleaning, assignment and pruning I describe above. As part of estimating a topic model, the *stm* package (Roberts et al. 2018) estimates a vector of probabilities for each word in the vocabulary, for each topic in each model. I use these values for each of the ambiguous topics I list above to calculate a correlation matrix, and then cluster the resulting correlations to identify the underlying relational structures in the model outputs.

Figure 3.1 below presents the end results this analysis. Each of the topics itemized in Table 3.2 occupies both a row and a column, so that the diagonal represents the perfect self-correlations of each topic with itself. The circles in the figure represent statistically significant correlations at the intersection of a given row topic and column topic, and are both shaded (darker means a higher value) and filled in with blue (more of the pie colored blue means a higher value) to indicate the amount of correlation. I removed all correlations with $p < 0.01$, leaving empty squares for those cases. Finally, I use hierarchical clustering with first three clusters (outlined in orange dashed lines) and then nine (outlined in black) as subsets of the larger three.

The lack of statistically significant correlations across most of the squares indicates that topics clustered fairly cleanly groups of two or three versions of a topic, which means that calculating similarities while taking the entire, 12,000+ word vocabulary into account provided a clearer picture of the relationships between them than was apparent in the top 15 or even 25 most probable terms in each.[2]

[2]That some were clusters of two and others clusters of three topics is consistent with my 2/3 rule for concordance, and is not problematic

Figure 3.1: Cross-Model, Mid-Level Topic Correlations



The hierarchical clustering provided some answers when three clusters (outlined in the orange dashed line) were imposed, and further clarified matters with the use of more, and more tightly defined, clusters (outlined in black). I used the incidence of smaller clusters - with their higher internal correlations - to identify the concordance labels that unified the content across their components. Beginning with the top left of the figure, the model identified two smaller, related clusters for topics describing block production using the Proof-of-Work method or mining and another for block production using other methods. Surprisingly, these two clusters had no significant relationship with the three topics I labeled *Cryptocurrency*, indicating that the conflation of special-purpose blockchains for cryptocurrencies with the currencies themselves in the Top-Level topics was not borne out at the Mid-Level. This also implies that **Blockchain Production** as a Top-Level topic should capture all Mid-Level aspects of these blockchains, rather than including them in **Cryptocurrency** as a Top-Level category.

Moving down the diagonal in Figure 3.1 to the next cluster, the model confirms the relationship between *Cryptocurrency* and *Crypto Payments* by placing them next to each other. That said, it also indicates surprisingly little overlap between them given the general lack of correlations between individual topics across these smaller clusters. I leave the process of further disambiguating these to the following chapter. For the moment, I note that I included a fourth potential payment topic (Topic 32, from Mid-Level model 2) given its ambiguous shared emphasis on payments, banking and financial inclusion. The correlation and clustering analyses both reflect these similarities between Topic 32 and the other Crypto Payments topics, but ultimately separated it from the three I use as concordance topics.

The final group of concordance topics all involved the mechanisms and technologies through which blockchain serves as general-purpose infrastructure for other applications. The topics here for the most part sorted neatly into two- or three-topic clusters capturing aspects of the technology including *Smart contracts*, *APIs and developer tools* and the widespread use by other applications of *Ethereum as infrastructure*. The more challenging cluster was the five topics that more generally capture the nature of blockchains in general as infrastructure. A separate analysis (not included here but available on request) using a different correlation measure confirmed all of the output in Figure 3.1 with the exception of *Smart contracts*, which it combined into the larger infrastructure cluster while separating two of the five topics clustered here under infrastructure into their own group, which I gave the concordance label *dApp Platforms*. I take this last concordance topic, along with those labeled in Figure 3.1, as the disambiguated Mid-Level concordance topics for those itemized in Table 3.2.

3.4 Preliminary Findings

The primary output of the process I describe above is a three-layer taxonomy of blockchain implementations. The taxonomy has 23 top-level categories, 48 mid-level categories, and anywhere from three to 12 detail-level categories under each of the mid-level categories. The depth and breadth of the taxonomy provide a reasonable early basis for defining blockchain technology as a GPT given the combination of new categories (e.g. Blockchain Production) and a wide range of pre-existing domains (e.g. Physical Production).

Table 3.3 adds the Top-Level categories from my taxonomy to those from the expert taxonomies used in economic studies of GPTs that I first in Table 2.1. In keeping with the nature of GPTs, as well as the established precedent of expert taxonomies, the top-level categories in my taxonomy span roughly similar territory as those of earlier taxonomies, even as the addition of new technologies (e.g. blockchains and cryptocurrency) necessitate new categories beyond those envisioned in the incumbent taxonomies.

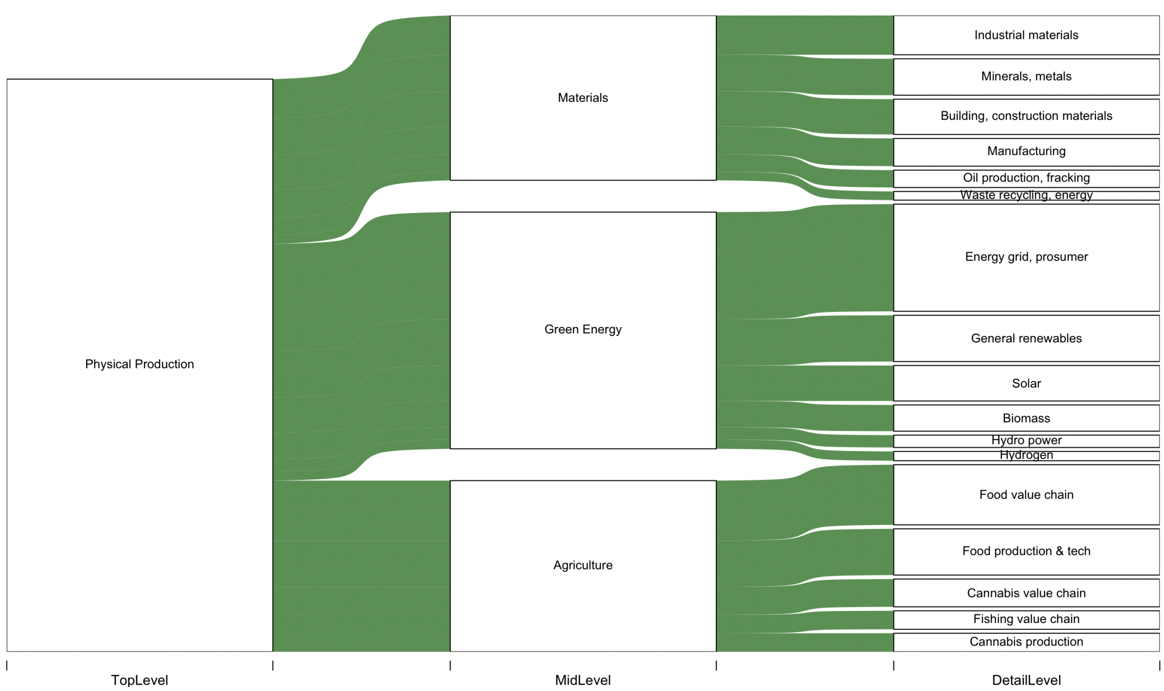
Table 3.3: Top-Level Categories in Context

CPC Technology Taxonomy *Patents*	**Blockchain Taxonomy** *Implementations*	**NAICS Production Taxonomy** *Industries*
Human Necessities	**Blockchain Production**	Agriculture, Forestry, Fishing, Hunting
Performing Operations; Transporting	**Blockchain as Infrastructure**	Mining, Quarrying, Oil & Gas Extraction
Chemistry; Metallurgy	**Cryptocurrency**	Utilities
Textiles; Paper	**Crypto Payments**	Construction
Fixed Constructions	**Investment**	Manufacturing
Mechanical Engineering; Lighting; Heating; Weapons; Blasting	**Trading**	Wholesale Trade
Physics	**Lending, Money, Banking**	Retail Trade
Electricity	**Physical Production**	Transportation and Warehousing
Various Technology	**Gaming, Esports, Gambling**	Information
	Art, Entertainment	Finance and Insurance
	Content, Advertising	Real Estate, Rental, Leasing
	Retail, Marketing	Professional, Scientific, and Technical Services
	Real Estate, Transport, Travel	Management of Companies and Enterprises
	Data Generation, Processing, Trading	Admin., Support, Waste Mgmt. & Remediation
	Distributed Computation, Devices	Educational Services
	Data Management, Storage	Health Care and Social Assistance
	Medical, Health	Arts, Entertainment, and Recreation
	Mobile Apps	Accommodation and Food Services
	Donations, Social Impact	Other Services (ex Public Admin.)
	Claims Evaluation	Public Administration
	Verification, Certification	
	Proof, Encryption	
	Governance	

Developing this taxonomy is a necessary first step to answering the dual challenges of identifying a GPT – identifying the generality of the technology's components, and the breadth of its application – but is not in itself sufficient. The distinctions between the nature of these two aspects is perhaps clearest in the two dominant patterns I found while constructing the taxonomy in I present at the end of this chapter.

The first of these appeared consistently across existing industries and technologies, where the challenge of identifying the breadth of the new technology's application was the primary challenge. In the vast majority of these categories, both the topic/categories and their hierarchical cooccurrences across levels were largely consistent across and within models. For example, projects assigned to the top-level Physical Production category were most often also assigned to mid-level categories for three overlapping types of physical production: agriculture, green energy and materials. While there was some overlap between the detail-level assignments given the overlapping nature of the mid-level categories, this overlap was minimal. Figure 3.2 below provides a stylized representation of the relationships between these categories.

Figure 3.2: Physical Production Sub-Taxonomy



While this outcome is an encouraging first step, it doesn't quite close the distance to answering the second of the dual challenges in the context of identifying a GPT. The taxonomy provides a baseline framework for determining the breadth of applications of blockchain technology, but leaves unanswered the deeper question of how the technology is used across these domains. I return to this question in Chapter 5.

In the meantime, I turn to the second pattern in the data, namely the presence of a set of topics involving the core components of blockchain technology that were far more promiscuous in their mingling, cooccurring with nearly every other higher-order topic when analyzed (i.e. Mid-Level topics that cooccurred with nearly every Top-Level topic, and Detail-Level topics that cooccurred with nearly every other Mid-Level topic). From a taxonomic perspective, these topics were aberrations that threatened the analysis through their categorical ambiguity. From the perspective of digital transformation via blockchain, however, they proved far more interesting. These topics included both most of the blockchain- and cryptocurrency-related topics, as well as others at the Mid- and Detail-Levels that clustered around a series of innovations enabled by the technology that are used widely across the ecology. As a result, rather than being problematic exceptions to the categories in the taxonomy, these frequently appearing topics provide evidence of the infrastructural and general-purpose nature of blockchain technology. In other words, they provide the basis for answering the first challenge of identifying a GPT, namely identifying both the components of the technology and their generality of purpose. They make clear which specific aspects of the technology are infrastructural as well as the nature of the infrastructure they provide. The very commonality of these topics leads me to term them general-purpose topics, and their importance leads me to devote the next chapter to articulating them more clearly before going on to explore their enabling role across categories in the taxonomy.

The remainder of this chapter presents the taxonomy I developed in the analysis described above.

Table 3.4: Detailed Taxonomy of Blockchain Implementations

Top-Level Categories	Mid-Level Categories	*Detail-Level Categories*
Blockchain Production	Block Production	*PoW mining* *PoS staking* *DPoS block production* *Master, supernodes*
	Mining Hardware	*Mining chips (ASIC, GPUs)* *Mining servers* *Mining/datacenter infrastructure* *Mining and renewable energy*
	Scaling, Throughput	*Sharding* *Interoperability* *Sidechains* *Childchains* *Payment channels* *Fate channels* *Ethereum layer 2 (e.g. Plasma)* *Bitcoin layer 2 (e.g. RSK)*
	Non-PoW Consensus Protocols	*Sharded consensus* *Casper, finalization* *Tendermint, BFT consensus* *Verifiably Random Function (VRF)* *Supermajority, replica consensus* *Dfinity, notarization proof* *Cardano consensus* *Wireless consensus* *FBA consensus (Stellar)* *Maidsafe consensus, PoR* *POET consensus*
Blockchain as Infrastructure	API, Developer Tools	*DApp development* *App developer tools* *Smart contract libraries (e.g. Zeppelin)* *Development infrastructure (e.g. IPFS)* *API tools, IDE* *Microservices, API architecture* *General developer tools* *Distributed software development* *Open source software development* *Devops and automation*
	Enterprise	*Supply chain*

Table 3.4: (continued)

Top-Level Categories	Mid-Level Categories	*Detail-Level Categories*
		Enterprise Resource Planning (ERP) *Software-as-a-service (SAAS)* *Factoring* *Accounting* *Trade documentation*
	Permissioned, Hyperledger	*Hyperledger blockchains* *(e.g. Fabric, Sawtooth, Iroha)* *Consortium blockchain* *Hyperledger modules and tools* *(e.g. Besu)*
Data	Data and Markets	*Personal data* *Business data*
	AI, Machine Learning	*Artificial intelligence (AI)* *Machine learning (ML)* *Neural network* *ML tournament* *Data science, predictions* *Robotics* *AI trading, investment*
	Distributed Computation	*Fog computing* *Enterprise cloud infrastructure* *(e.g. AWS, Oracle)* *Distributed cloud infrastructure* *Distributed web, data hosting* *Content Delivery Networks*
	Distributed Data Storage	*Distributed data storage platforms* *Distributed storage infrastructure* *(e.g. IPFS)* *Filesharing, torrent networks*
	Networked Devices	*IOT sensors* *WiFi hotspot, hosting* *Mesh network, devices* *Mobile, 5G data* *Location beacon*
Asset Finance	Investment	*Investment fund* *Portfolio management*

Table 3.4: (continued)

Top-Level Categories	Mid-Level Categories	*Detail-Level Categories*
		Index fund management *Wealth management* *Pension, retirement* *Price index* *Quantitative portfolio management*
	Digital Assets	*Tokenization and digital rights* *Underwriting digital assets* *Security token issuance* *NFT issuance* *Token storage, custody*
	Trading	*Trading strategies* *Trading platforms* *Copy trading* *Signal trading* *Momentum trading* *Arbitrage trading* *Trading bots*
	Trading Infrastructure	*Centralized crytpo exchanges* *Decentralized crypto exchanges/DEX* *DEX relayers* *Automated market making* *Dark trading pools* *Traditional securities brokerage* *Trading engine, orderbook* *Margin trading* *Futures trading*
Money Finance	Cryptocurrency	*Cryptocurrency* *Gold-backed cryptocurrency* *Masternode networks* *Privacy transactions* *Zero-Knowledge transactions*
	Stablecoin	*Algorithmic stablecoin* *Pegged stablecoin (e.g. gold, USD)* *Collateralized stablecoin* *Lending-based stablecoin*
	Crypto Payments	*P2P payments* *Payment, debit, atm cards* *Wallet*

Table 3.4: (continued)

Top-Level Categories	Mid-Level Categories	*Detail-Level Categories*
		Cryptobanking
		Mobile banking and inclusion
		Remittances, crossborder (retail)
		Remittances, crossborder (business)
		Payment infrastructure (cryptocurrency)
		Merchant, retail payments
		Cannabis payments
		Porn, webcam payments
		Travel payments
	Lending, Credit	*P2P lending platform*
		Credit scores
		Borrower guarantees
		Mortgage lending
		Loan securitization
		Debt markets, derivatives
		Microfinance, lending circles
	Precious Metals	*Precious metals*
		Jewelry
	Banking Infrastructure	*Central Bank Digital Currency (CBDC)*
		Central Bank settlement systems/RTGS
		Interbank payments, infrastructure
		Regulatory compliance, KYC/AML
Art & Entertainment	Art, Collectibles	*Art market*
		Museum, art collection
		Digital art
		Antique market
		Wine, spirits
		Exotic automobiles
	Entertainment	*Music production*
		Music streaming
		Music performance
		Film financing, production
		TV financing, production
		Film, tv streaming
		Live events, tickets
		IP, royalties
Content &	Advertising	*Online advertising*

Table 3.4: (continued)

Top-Level Categories	Mid-Level Categories	*Detail-Level Categories*
Advertising		*Programmatic advertising* *Affiliate marketing* *Screen, display advertising* *Social network, social media adv* *Influencer marketing* *Advertising attribution, fraud*
	Content	*Content creation, distribution* *Content monetization* *Stock photography* *Book publishing* *News, professional media* *Korean television*
	Video	*Video streaming* *Livestreaming* *Pornography, webcam* *Streaming infrastructure*
Gaming & Gambling	Game Development	*Game development platform* *Ingame goods design* *Ingame goods NFTs, trading* *Ingame economy, currency design* *Game entrepreneurship* *Virtual worlds*
	Gameplay	*Gaming* *Collectibles, creatures, NFTs* *Role playing games*
	Gambling, Betting	*Online casino* *Poker* *Raffle* *Slots* *Lottery* *Dice betting* *Gambling infrastructure* *(e.g. fate channels, rng)*
	Sports, Esports	*Fantasy sports* *Esports tournament* *Sports, esports betting* *Sports celebrities* *Freelance talent*

Table 3.4: (continued)

Top-Level Categories	Mid-Level Categories	*Detail-Level Categories*
		Streaming, livestreaming events
		Soccer
		Tennis
	Mixed Reality	*Virtual reality*
		Augmented reality
		VR, AR hardware
		3D, immersive environment
		Virtual land
Physical Production	Agriculture	*Fruit, vegetables, livestock value chain*
		Cannabis value chain
		Coffee value chain
		Fishing value chain
		Food supply chain, traceability
		Organic agriculture
		Grocery store, distribution
	Materials Production	*Building materials*
		Home construction
		Geological mining, exploration
		Oil exploration
		Oil drilling
		Dioxide production
		Graphene production
		Steel production
		Manufacturing
		3D manufacturing
	Green Energy	*Solar energy production*
		Solar energy microgrids, prosumer
		Biomass energy
		Hydro energy
		Hydrogen fuel
	Environment	*Climate mitigation*
		Carbon credits
		Carbon markets
		Water treatment
		Tree planting
		Eco sensors
Transportation &	Shipping & Logistics	*Maritime shipping*

Table 3.4: (continued)

Top-Level Categories	Mid-Level Categories	*Detail-Level Categories*
Real Estate		*Freight carrier, trucking*
		Cargo transport and logistics
		Parcel delivery
		Invoice, trade financing
	Automobiles	*Automobile sales, rental*
		Ridesharing
		Taxi
		Parking
		Elect. Vehicle Charging
		Automobile provenance
		Automobile data
	Travel	*Travel booking*
		Hotel booking
		Sharing economy rental
		Adventure travel
	Real Estate	*Real estate listings, data*
		Real estate agents
		Property rental marketplace
		Investment property
		Land title registry
Claims & Evaluation	Education	*College education*
		Skill development
		Certification, verification
		Tutoring
		Elearning
	Employment	*Gig/freelance work, marketplace*
		Dispute moderation, arbitration
		Crowdsourcing
		Contractor reviews
		Recruiting, traditional employment
	Insurance	*Insurance policy, premiums (check this)*
		Travel, parametric insurance
		Mutual, P2P insurance
		Insurance syndicate
		Reinsurance
	Reputation, Ratings, Evaluation	*Portable reputation score*

Table 3.4: (continued)

Top-Level Categories	Mid-Level Categories	*Detail-Level Categories*
		Ratings, reviews *Attestations, peer endorsement* *Moderation* *Tribunal*
Verification & Governance	Identity	*Self-sovereign identity* *Biometric identity* *Credential Attestation*
	Market, Arbitration	*Prediction markets* *Legal, jury dispute resolution* *Arbitration*
	Governance	*DAO platforms* *Voting* *Governance protocols*

Chapter 4

Blockchain's Generality of Purpose

I suggest in the prior chapter that the topics that intersect most broadly with other topics in their cooccurrences are indicators of aspects of the technology that carry its generality of purpose. From the perspective of digital transformation, it would be reasonable to expect these to be the technology-specific mechanisms that enable digital transformation. Analyzing their scope and interactions should thus provide an entry point for a deeper analysis of how blockchain itself works as a general purpose technology, and in particular its role as a potential engine of digital transformation.

Doing so requires further analysis to identify these topics consistently, as well as the functions that underwrite their generality of purpose. I draw on the output of the modeling I describe in the prior chapter, though for this chapter I work primarily from interpretation of the semantic content of the topics I identify. I focus this interpretive process on identifying affinities and linkages across topics (and across levels) this content reveals. Doing so allows me to focus on the practical aspects of blockchain as a technology-in-use as identified in the topics, and to use this to develop insights into the specific nature of digital transformation enabled by blockchain.

4.1 Approach

The first step in identifying general purpose topics is to distinguish them from topics that just have many projects assigned to them or appear at the top of the three-level hierarchy. While the latter do have aspects of being general, they do not necessarily rise to having generality of purpose, which is to say that they don't necessarily represent aspects of the technology that meet the definition of a general purpose technology. I build on prior work on general purpose technologies to define the criteria for identifying one in my data. I define general purpose topics relationally in the context of other topics and the linkages they create across levels. More specifically, I define them as Mid-Level (Detail-Level) topics that cooccur frequently with many Top-Level (Mid-Level) topics, and whose cooccurrences are more frequently substantial than the other topics at a given level. The former requirement captures the breadth of use of general purpose technologies, while the latter establishes a threshold for relevance.

This leaves the question of specifying the thresholds for "frequent" and "substantial". I define "substantial" as those cooccurrences with at least five projects assigned to them for Mid-Level topics, and at least three for Detail-Level topics, with these numbers drawn from my own analysis of relative frequencies at each level. I take a similarly data-driven approach to defining the threshold for frequency by looking at each set of co-occurrences (e.g. those between topics in Top-Level Model 2 and those in Mid-Level Model 3). I plotted histograms of the frequency of substantive cooccurrences across all of the pairings, and then visually identified natural breakpoints in these charts.[1]

Doing this for Mid-Level cooccurrences with Top-Level topics, and then for Detail-Level cooccurrences with Mid-Level topics, produced a list of 91 topics spanning both levels. I list these topics in Table 4.1 below, with the Mid-Level general purpose topics on the left and Detail-Level general purpose topics to the right.

While this table implies many possibilities for interpretation, it requires some structure to make the findings legible. Building on Chapter 2, I proceed from the assumption that topics

[1] Analysis available on request.

Table 4.1: Mid- and Detail-Level General Purpose Topics

MID-Level GPTopics	Detail-Level GPTopics	
API, DEV TOOLS	ai, ml, deep learning	messaging mech
BLOCKCHAIN AS INFRA	algorithm	mining farm, rig
BUSINESS PROJECTIONS	app dev tools	ml model, analytics
COIN VALUE MECHANISMS	basic income, passive	mobile app, appstore
CROWDFUNDING	bitcoin	network attack
CRYPTO PAYMENTS	block production	networked servers
CRYPTOCURRENCY	bounty	nonprofit org
DATA	business affordances	options
ECOSYSTEM	business templates	os protocol layer
ENTERPRISE	centr, decentr exchange	P2P network
ETH SMART CONTRACT, ORACLE	chain structure	payment gateway
ETHEREUM AS INFRA	coin mechanisms	platform affordances
INVESTMENT	community	pow mining
MARKETPLACE	content creat, distribut	pow, pos
POOLS, REWARDS, INCENTIVES	crowdfund campaign	project development
PROJECT DEVELOPMENT	crypto payments	provider mktplace econ
TOKENIZATION, DIGITAL ASSETS	currency, digital currency, value	RE listings
TOKENS	customer loyalty prgm	reserve, kyber
USER AFFORDANCES	dapp development	rewards, incentives
	data storage	security, password, pk
	device, hw, remote	selfsov identity
	digital ownership	service provision
	distr storage, fileshare	smartcontract semantics
	distrib database	staking mechanism
	ecommerce marketplace	startups, entrep
	economic, social system	tokenization, digital rights
	ecosystem development	tokenization, underwriting
	enterprise affordances	tokens
	enterprise blockchain	tx code
	ethereum smartcontracts	tx code.1
	fee pool	tx confirmation P2P
	foundation governance	user affordances
	gaming, ingame virtual curr	video streaming
	inperson events	voting, decision
	investment fund.1	wallet
	marketplace tx	
	merchant payments infra	

represent components of larger functional clusters, and look to identify these clusters based on my data and model output. The widespread prevalence of these topics, in terms of frequency of cooccurrences with so many other topics, blunts the effectiveness of the approach I developed in the prior chapter. Instead, I look to identify the nature and contribution of functional clusters interpretively through analysis of the content of the topics themselves. This process involves both the interpretive analysis of topic content as well as interpretive disambiguation of remaining areas of overlap not addressed by the computational disambiguation in the prior chapter.

Through this process I identify the four primary dimensions of the generality of purpose of the technology whose implementations comprise my data set. The first of these is *computational infrastructure*, or the creation and provisioning of generative digital infrastructure. The second is *programmable money*, or the creation of cryptocurrency and other new forms of digital money (such as central bank digital currencies) that operate inside and outside of traditional legal tender and are created by code. The third is the creation of digital assets and their representations, which I term *tokenization*. The fourth and final dimension is *sociotechnical organization*. I describe the components of and rationale for each in the sections that follow.

4.2 Computational Infrastructure

The core concept of "blockchain" as digital infrastructure was first put into action with the launch of Ethereum's beta or Frontier version of its network in 2015. This infrastructure is designed to be both specific and coherent in its inner functioning and open and generative in its adaptability for use by third parties. The specific nature of the infrastructure is computational, meaning that it derives all of its desirable features (including immutable records supporting property rights, conditional execution of code for third-party applications, and cryptographically secure data storage) through large-scale computation by distributed networks of participants willing to contribute computational power to the network. This form of infrastructure is closely related to that of cloud computing in function, but distinct in its avoidance of centralization of resources and in its potential to use project-specific money in an incentive structure for participants.

I build on this notion of computational infrastructure to define the first cluster of general purpose topics, which appear in Table 4.2. The table includes representative content or terms from each of the Mid-Level general purpose topics I've selected as part of the cluster, as well as those from a subset of the related Detail-Level general purpose topics I've selected; the remainder of the associated Detail-Level topics appear at the end of the table.[2]

Blockchain as Infrastructure is the most general of these in its inclusion of terms describing applications, decentralized applications (or dApps) and development, along with descriptive verbs

[2]The terms capturing the content of these topics in this and other tables in this chapter reflect the data cleaning I describe in the prior chapter, particularly the joining of terms such as "smart contract" into a single word.

Table 4.2: Computational Infrastructure

Mid-Level Topics & Content†
Data *(Model 2, 86)* data; personaldata; bigdata; datasource; dataexchange; analytics; dataset; insight; gdpr; provider; privacy; collected; ocean; collect; datastorage; analyze; userdata; analysis; sharing; collection
Developer Tools *(Model 3, 62)* module; api, interface; template; integration; backend; component; developer; application; microservice; tool; workflow; framework; custom; plugin; frontend; software; architecture, sdk; javascript
Blockchain as Infrastructure *(Model 2, 39)* dapp; layer; developer; application; module; api; neo; blockchain; template; framework; sdk; architecture; blockchainnetwork; eos; interface; programming; decentralized; publicblockchain; plugin; blockchainapplication
Enterprise *(Model 2, 50)* supplychain; enterprise; accounting; invoice; privateblockchain; business; erp; immutable; blockchainsolution; automation; audit; transparency; record; management; immutability; tracking; workflow; blockchaintechnology; traceability; sme
Ethereum as Infrastructure *(Model 3, 6)* evm; gas; vm; script; stack; opcode; instruction; solver; truebit; relay; computation; nonce; execution; turingcomplete; bytecode; scripting; utxo; byte; computational; state
Ethereum Smart Contracts *(Model 3, 14)* contract; oracle; smartcontract; offchain; onchain; smart; relayer; derivative; logic; execute; agreement; chainlink; call; template; ricardian; external; contractual; erc20token; sc; counterparty
Selected Detail-Level Topics and Content†
Dapp Development (Model 1, 540) dapp; line; deploy; core; developer; framework; tokeneconomy; cycle; userexperience; authority; dsp; digitaltoken; building; usage; store; decentralized; portal; library; native; interact
Data Storage (Model 2, 235) data stored; information; layer; realtime; integrity; processing; verify; access; type; datastorage; sharing; collection; source; system; request; encrypt; collect; architecture; mechanism
Distributed Database (Model 2, 467) database; mongodb; query; relational; characteristic; sql; stored; enterprise; deployment; core; administrator; control; scenario; centralized; integrity; scale; consortium; attribute; operational; single
Distributed Storage (Model 2, 514) file; storage; datastorage; storing; stored; copy; download; upload; space; store; storj; encrypt; host; decentralized; sia; uploaded; encryption; drive; disk; size
Protocol Layer (Model 3, 27) protocol; decentralized; thirdparty; built; layer; enable; implementation; usecase; offchain; opensource; mechanism; directly; top; onchain; interface; value; centralized; function; open; entity; create
Associated Detail-Level Topics
App Developer Tools; Ethereum Smart Contracts; Smart Contract Semantics; Business Templates; Business Affordances; Enterprise Blockchain; Enterprise Affordances

† Topic content presents the 20 words with the highest probability according to the FREX algorithm. The words are taken from the model indicated, and appear consistently across at least two of the three Mid-Level models.

(e.g. create, support). This Mid-Level topic's content includes the term protocol, a typical building block of both blockchains and the applications they enable, and a Detail-Level topic in its own right (*protocol layer*). The topic's content also includes other tools, primarily APIs (application programming interfaces) and code modules. Both APIs and developer tools are further detailed in the content of the *API, Developer Tools* Mid-Level topic, and are also captured in Detail-Level topics including *dapp development*, *app development tools*.

It is unusual but apt that Ethereum is the subject of two different Mid-Level topics, since it is the only individual project for which this occurs. The two topics each capture an aspect of the project's technological innovations, each of which became a template for the builders of later blockchains projects. The first innovation was to use the aggregated computational power of its peer-to-peer network as a general-purpose computational engine, rather than solely to secure its network as Bitcoin does (Buterin 2013; Wood 2014). This conception of the network providing an Ethereum virtual machine (or "EVM", a nod to Java's existing virtual machine or "VM" technology) for computation is reflected in the **Ethereum as Infrastructure** topic. The second Mid-Level Ethereum topic captures the project's other major innovation: smart contracts. This technology was conceived two decades before (Szabo 1997) but first put into practice at scale by Ethereum as a means of allowing independent applications to connect to and draw on a blockchain's computational processing and recording of transactions.[3] The dominant role of Ethereum after pioneering this model of computation and smart contracts is reflected in the fact that Ethereum hosted just over 80% of dApps in late February 2021.[4]

The third Mid-Level component of this cluster is *Enterprise*, or the development and use of blockchains for large organizations. The content of this Mid-Level topic captures both blockchain as a technology as well as terms describing focused uses in information processing and recording (e.g. process, data, information, record, document). It also captures some of the most frequently implemented uses for blockchain systems for business, such as supply chain or risk management ("ERP"). These are, however, far from the only applications for the technology when used by businesses or enterprises, as evidenced by the general-purpose nature of related Detail-Level topics including *business templates*, *business affordances*, *enterprise blockchain* and *enterprise affordances*.

These three visions of blockchain as infrastructure intersect at multiple points with the **Data** topic, particularly in its components having to do with distributed data storage and managing distributed databases. Blockchain ledgers themselves are the result of both distributed

[3]For a more thorough discussion of smart contracts, see Chapter 2.
[4]https://web.archive.org/web/20210310152408/https://www.stateofthedapps.com/stats

computation and distributed storage performed by networks of servers. Many of the same mechanisms are used by these ledgers and by projects that focus more closely on the storage component rather than producing a ledger. Distributed databases are also a core concern for large corporate and state actors whose interests are captured by the *Enterprise* topic at the Mid-Level, and are central to the use cases listed there (e.g. ERP). Storage is also a primary concern for the construction and running of Dapps, with many relying on distributed data storage projects such as IPFS or Storj to record and store their users' data on a separate information layer from whichever blockchain they use for transactions.

4.3 Programmable Money

One of the most commonly used phrases among practitioners, which is to say developers and owners of digital coins, refers to these media as "programmable money," meaning money whose parameters and ongoing management are based in code rather than the authority of state institutions. While the ability to affix conditions to money is not new (Koning 2020), the combination of programmability and the possibility of shifting management and authority from the state to distributed systems is a unique contribution of blockchain technology (Lee 2021).

The ability to create programmable money has expanded the domain of digital coins well beyond cryptocurrencies, though this expansion can be difficult to see given the dominant role of Bitcoin in both the popular press and academic research. Just as Bitcoin often stands in for the entire space, the term cryptocurrency is often used for all traded digital coins. This situation also results from the fact that many kinds of digital coins trade on exchanges, leading to an implicit flattening of distinctions. This conceptual ambiguity about the nature of digital money is not surprising given the wide variety of meanings already attached to government-issued currencies, let alone the welter of alternatives to legal tender that predated cryptocurrency (Zelizer 1997). The creation of programmable money makes it possible to align multiple uses with multiple media more directly than is the case with state currencies, which in turn necessitates a more detailed approach to classification than simply labeling everything "cryptocurrency."

To that end, I draw together the Mid-Level and Detail-Level topics I associate with aspects of programmable money in Table 4.3 below. I include representative content of all of the Mid-Level topics as well as selected Detail-Level topics as part of my analysis.

Table 4.3: Programmable Money

Mid-Level Topics & Content†
Coin Supply Management *(Model 3, 16)* coin; coinholder; mint; frozen; progressive; forever; burn; cryptocoin; failed; denomination; freeze; dead; king; currency; altcoin; circulation; bill; initial; holdings
Cryptocurrency *(Model 2, 58)* cryptocurrency; digitalcurrency; bitcoin; blockchaintechnology; currency; satoshi; peertopeer; litecoin; altcoin; revolution; money; world; paymentsystem; nakamoto; ripple; revolutionary; conventional; marketcapitalization; financialtransaction; cryptocurrencymarket
Crypto Payments *(Model 1, 23)* card; remittance; debitcard; creditcard; atm; banking; fiat; merchant; visa; paymentsystem; payment; crypto; bank; fiatcurrency; cards; money; wallet; gateway; prepay; fintech
Monetary Policy *(Model 3, 59)* monetary; reserve; inflation; currency; stability; economy; volatility; centralbank; stable; interestrate; economic; velocity; gdp; supply; exchangerate; ratio; money; demand; marketcap; rate
Tokens *(Model 1, 24)* token; tokenholder; numberoftoken; erc20; burn; airdrop; tokenvalue; utilitytoken; tokenprice; amountoftoken; erc20token; tokendistribution; utility; tokenuse; platform; tokensupply; tokenbased; ethereumblockchain; tokenize; value
Selected Detail-Level Topics and Content†
Coin Generation and Distribution (Model 1, 820) coin; second; coinholder; start; every; cryptocurrency; hold; per; send; distribution; introduction; unique; amount; get; day; feature; early; many; supply; premine
Cryptocurrency Payments (Model 1, 102) cryptocurrency; fiat; crypto; fiatcurrency; cryptocurrencymarket; cryptocurrencyexchange; money; payment; accept; exchange; buy; world; store; bank; purchase; volatility; growth; popular; everyday; launch
Digital Currency (Model 2, 188) currency; money; digitalcurrency; stable; monetary; economy; dollar; local; fiatcurrency; exchangerate; issued; global; value; economic; usd; circulation; cash; national; issuance; medium
Exchanges (Model 1, 323) exchange; dex; trading; centralized; cryptocurrencyexchange; volume; trade; withdrawal; crypto; listed; fee; deposit; exchanged; decentralized; listing; buy; stock; fund; withdraw; sell
Other Associated Detail-Level Topics
Bitcoin; Payment Gateway; P2P Transaction; Wallet; Tokens

† Topic content presents the 20 words with the highest probability according to the FREX algorithm. The words are taken from the model indicated, and appear consistently across at least two of the three Mid-Level models.

I begin from an obvious place, by tracing the appearance of the terms "currency" and "money." Among the Mid-Level topics, these terms are part of the significant overlap between the **Cryptocurrency** and **Crypto Payments** topics, which are even more closely connected by their shared emphasis on payments and transactions. The terms also appear in the **Monetary Policy** topic, in the context of traditional state economic and policy terms typically discussed in the context of central banks and state money supply, though this emphasis on state institutions would seem to make it a poor fit with the digital topics. By contrast, the **Coin Supply Management** topic addresses also currency supply but solely in the context of blockchain-based cryptocurrencies and does not include the term "money" among its highest-probability words. The implication of these topics is that cryptocurrency is a form of digital currency used for payments, though the distinction between cryptocurrency and traditional currency is only partially defined.

The distinction becomes clearer when considering the Detail-Level topics, as does the generality of purpose. The *Digital Currency* topic captures aspects of both the **Cryptocurrency** and **Monetary Policy** Mid-Level topics in its focus on both currency, exchange rates and stability. In doing so, it links digitally generated and exchanged currency with aspects of the classic definition of money from economics, namely serving as a medium of exchange ("cash") and a store of value ("stable," "value"). The integral connection between cryptocurrency and payments is perhaps clearest in their merging at the Detail Level into the single topic *Cryptocurrency Payments*, which manifests as a merging of the two Mid-Level topics and indicates that – of the varieties of digital currencies – cryptocurrencies are used primarily for payments. This leads me to define cryptocurrency as *digitally generated, secured and transacted currency used for peer-to-peer payments*, and crypto payments as *the digital infrastructure and user-facing application for peer-to-peer payments using cryptocurrency*. The latter definition is borne out in the content of the Detail-Level topics *Payment Gateway, P2P Transaction and Wallet*, each of which articulates an aspect of payments using cryptocurrency.

I also group the Detail-Level topic *Exchanges* in this cluster. While digital currency exchanges are part of the larger fintech grouping of categories (as I discuss in the following chapter), they are also central to the functioning of digital currencies as media. Like national currencies, can largely be *traded* on exchanges but can only be *used* in specific contexts. Unlike national currencies, there are now thousands of digital currencies, so that the ability to transfer from one digital currency to another (let alone to transfer into and out of digital currencies from state currencies) is a necessary condition for the entire ecology.

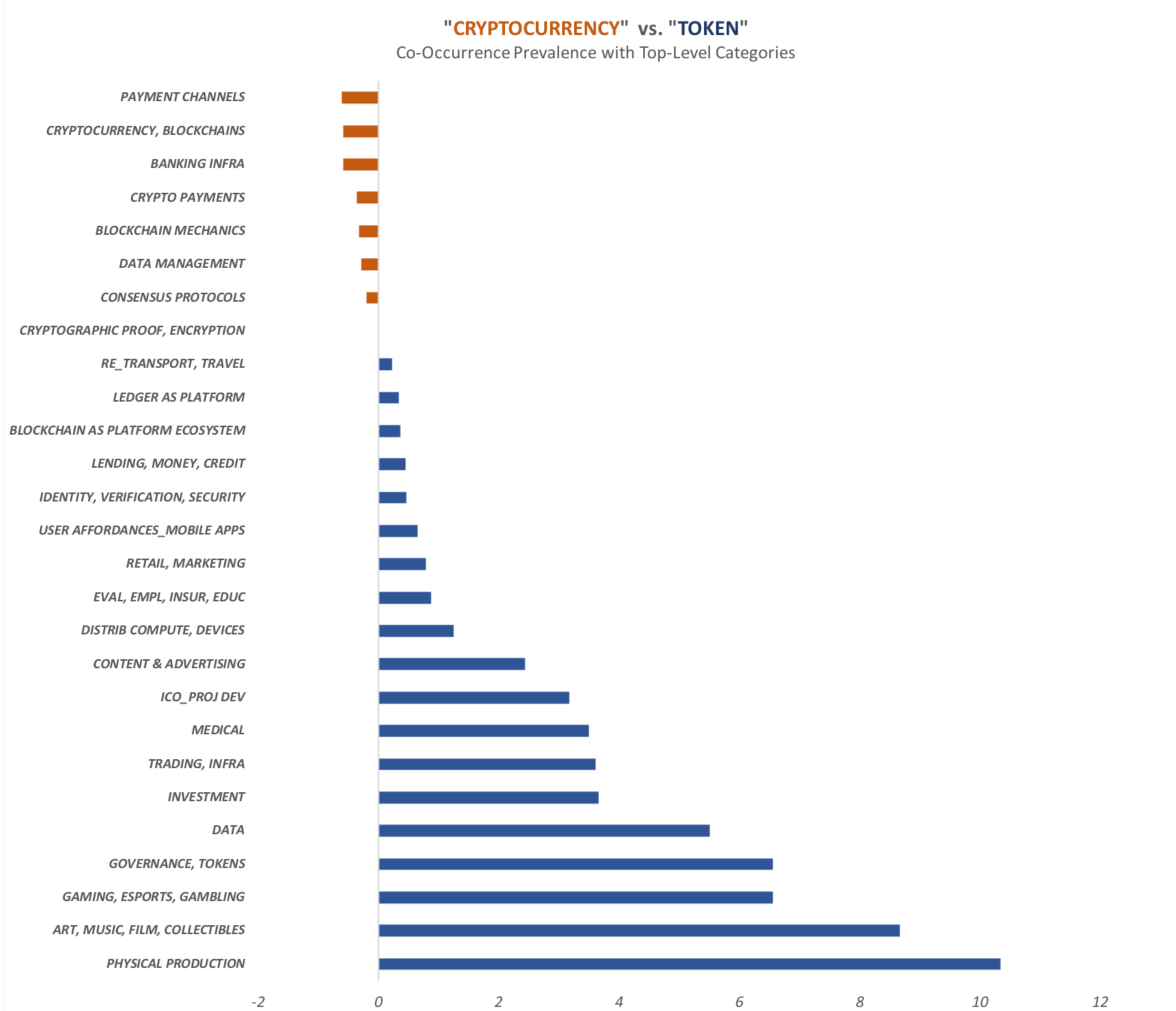
This leaves one final distinction to draw: the distinction between cryptocurrencies and tokens as forms of programmable money. The Mid-Level topics in Table 4.4 show a near-complete lack of overlap between these concepts, as do the Detail-Level topics, even though **Tokens** is among the most broadly assigned topics at both the Mid- and Detail-Levels. As a result, I turn from the content of the topics to look at them relationally, in the context of their cooccurrences with other topics. I start by looking upward, examining cooccurrences between each as a Mid-Level topic and the full spectrum of Top-Level topics. Put another way, I examine the way cryptocurrency and tokens (as Mid-Level topics) co-occur differently with each of the Top-Level topics. After a simple normalization of the frequencies, the pattern in Figure 4.1 emerges.[5] The bars that fall to the left of the vertical axis represent Top-Level topics that intersect with **Cryptocurrency** more frequently than with **Tokens**; the opposite is true for bars that fall to the right of that axis.

The analysis in Figure 4.1 indicates diverging meanings for cryptocurrency and tokens in the context of the Top-Level topics. The top end of the graph presents Top-Level topics that are relatively more associated with (meaning, that cooccur more frequently with) *Cryptocurrency* than with *Token*, while the bottom are the converse. The topics more closely associated with cryptocurrency in this analysis are those having to do with banking, payments and the mechanics of blockchain production. Tokens, however, are more closely associated with all of the other topics involving specific applications of the technology, as well as with the use of blockchains as an infrastructural layer.

This simple analysis points to one of the more novel aspects of blockchain as a general purpose technology, and in particular to the importance of programmable money as an innovation. In addition to providing the computational infrastructure for applications, blockchain technology also provides the ability to create special-purpose money that developers and entrepreneurs can link to specific objectives and modes of participation. In some cases, these tokens function much as in-game tokens do in virtual games, as awards to be earned and as the currency with which to purchase goods. In others, the tokens are attached to a range of decision and property rights that is

[5]The rebasing is the ratio of token cooccurrences to cryptocurrency cooccurrences, minus 1.

Figure 4.1: Disambiguating Cryptocurrency and Token: Looking to Top-Level Topics



steadily evolving, with many projects in my data set incorporating multiple tokens with different rights attached to them.

This development is part of a broader aspect of the technology I term Tokenization, and which I develop separately in the following section.

4.4 Tokenization

Tokenization is an umbrella term I use for two interlinked processes enabled by both computational infrastructure and the ability to create programmable money. Scholars working at the intersection of economic sociology, science and technology studies (STS) and political economy have identified these processes of creating “new instruments of knowledge, efficiency and value extraction” (Fourcade and Healy 2017, p. 10) as central mechanisms of digital transformation. In the context of blockchain technology, these processes are both enabled by and foster the creation of computational infrastructure and programmable money.

The first process is sometimes called “datafication” (Mejias and Couldry 2019; Sadowski 2019), and reflects the “data imperative” of the digital economy for companies to reorient their activities toward the widespread extraction, trading and processing of data (Fourcade and Healy 2017; Zuboff 2019). In contrast to the existing world of technology monopsonists such as Google and Facebook, however, the projects in my data set are oriented to opening this process of transformation to broader participation by everyone from individuals seeking to monetize their personal data to corporations seeking to source sophisticated machine learning algorithms to use on their own data on open markets on which the algorithms have been tested and rated. These uses and others are described in the Mid-Level *Data* topic in Table 4.4.

The second transformation has been theorized as “assetization” (Birch 2017; Birch and Muniesa 2020), or the process of turning both abstract and physical goods into a form of capital distinguished from commodities by its ability to generate rents. By this definition, datafication is an important component of assetization, but far from the only one. Within my data set, this process is realized through the creation of project- and, in some cases, use-specific tokens to which various rights are attached by design. Perhaps the most widely used form of these rights is property rights that are attached to digital assets, which are typically a claim on an object rather than the object itself that is rendered into data to become digital, and given a value through tokenization. In some cases, these tokens are unique and are intended to establish sole ownership of a good, such as an artwork. These tokens are called Non-Fungible Tokens or NFTs, and this form of ownership is implicit in the content of the Detail-Level *Digital Ownership* topic. In other cases, tokenization is used to create a digital asset intended to be traded on financial token markets and used in investment portfolios, as in the two other Detail-Level topics related to tokenization.

The process of tokenization (and the viability of programmable money more broadly)

Table 4.4: Tokenization

Mid-Level Topics & Content†
Data *(Model 2, 86)* data; personaldata; bigdata; datasource; dataexchange; analytics; dataset; insight; gdpr; provider; privacy; collected; ocean; collect; datastorage; analyze; userdata; analysis; sharing; collection
Investment *(Model 2, 78)* investor; investment; invest; portfolio; capital; fund; dividend; vc; equity; assetmanagement; manager; venture; profits; angel; hedgefund; return; exit; profit; startup; wealth
Tokenization, Digital Assets *(Model 2, 57)* asset; digitalasset; tokenization; unit; tokenize; colored; mc; issuer; witness; issuance; ownership; depository; parent; supernode; stable; attestation; fractional; definition; issued; issuing
Tokens *(Model 1, 24)* token; tokenholder; numberoftoken; erc20; burn; airdrop; tokenvalue; utilitytoken; tokenprice; amountoftoken; erc20token; tokendistribution; utility; tokenuse; platform; tokensupply; tokenbased; ethereumblockchain; tokenize; value
Selected Detail-Level Topics and Content†
Digital Ownership (Model 3, 12) digital; physical; ownership; work; potential; unique; image; view; problem; transferred; right; non; trusted; crypto; internet; create; original; copy; access; solve
Exchanges (Model 1, 323) exchange; dex; trading; centralized; cryptocurrencyexchange; volume; trade; withdrawal; crypto; listed; fee; deposit; exchanged; decentralized; listing; buy; stock; fund; withdraw; sell
Tokenization, Digital Assets (Model 3, 196) asset; digitalasset; ownership; transfer; traditional; class; right; issue; manage; financial; tokenize; value; type; management; trade; assetmanagement; transparent; process; issuance; provide
Tokenization, Underwriting (Model 3, 109) company; dedicated; share; investor; employee; revenue; investment; shareholder; target; holder; corporate; sell; private; financing; financial; equity; invest; right; partnership; securities
Token Reserves (Model 2, 24) reserve; kyber; rate; epoch; automated; parameter; trade; conversion; source; burn; operator; exchangerate; taker; master; implementation; onchain; price; enabled; maintainer; destination; decide
Other Associated Detail-Level Topics
Startups, Entrepreneurship; Investment Fund; AI, ML, Deep Learning; ML Analytics; Algorithm; Data Storage; Distributed Storage, Filesharing; Distributed Database

† Topic content presents the 20 words with the highest probability according to the FREX algorithm. The words are taken from the model indicated, and appear consistently across at least two of the three Mid-Level models.

fundamentally relies on the existence of liquid markets. This is in part for the reason I describe in the section above, as individuals can only use certain coins as users of a given project and thus need to exchange tokens when moving from one project to another. Another important reason is that the incentives and uses encoded into these tokens rely on their having a monetary value that changes with value of the overall project. For example, the Audius music project I described in the introductory chapter has a publicly traded token with the ticker AUDIO that music consumers and producers use as their medium of exchange for payments and incentives within the project. The AUDIO token is listed on several token exchanges, opening it to trading on those markets and thus the assessment of investors. In theory, an increase in the amount of music being offered and consumed on Audius should be recognized by external traders who would bid up the price of AUDIO accordingly on external markets, and thus make the incentives for participating within Audius (and thus being compensated in AUDIO) even stronger.

The central importance of external markets to the use of tokens is the less obvious driver of the rapid growth of token exchanges, investment platforms, funds and other financial infrastructure that otherwise appears as speculative froth. The creation of token investment funds has thus been both a major activity of individual projects (as I describe in the following chapter) as well as an important part of the infrastructure of the larger ecology. I thus include topics on investment and trading in this cluster, as well as one that describes tokenized mechanisms for maintaining liquidity (Token Reserves).

4.5 Sociotechnical Organization

The final set of general purpose topics are not in themselves taxonomic categories, but rather are forms of sociotechnical organization that are used universally across implementations. I include only topics describing aspects of organization used primarily within projects and present these in Table 4.5.

These include a series of Mid-Level topics that function as basic building blocks for creating implementations of blockchain technology. While each of these is the subject of significant attention in various organizational literatures, they are typically not considered jointly as a set of building blocks available to entrepreneurs. The subject is sufficiently rich that I address these in a separate chapter, and for the moment focus on aspects of the set that reflect organizational structures enabled by tokenization (**Crowdfunding**) and that enable computational infrastructure (**Pooled Rewards and Incentives**).

As a topic, **Crowdfunding** is both functionally and semantically tied to the building and development of projects (also a Mid-Level topic in this group), for which initial coin offerings or "ICOs" have been a major source of initial capital. The advent of both digital currencies and

Table 4.5: Sociotechnical Organization

Mid-Level Topics & Content†
Crowdfunding *(Model 2, 67)* crowdfunding; backer; project; ito; campaign; startup; fundraising; teammember; funding; milestone; entrepreneur; translation; raise; projectteam; english; tokenholder; team; completion; financing
Ecosystem *(Model 1, 26)* ecosystem; adoption; economy; foundation; vision; innovation; incentivize; sustainable; mass; drive; engagement; leverage; leveraging; accelerate; enabling; engage; community; align; adopter; utility
Marketplace *(Model 2, 60)* seller; buyer; marketplace; vendor; auction; bid; supplier; listing; escrow; invoice; goods; mediator; item; ebay; bidder; buying; sell; goodsandservice; buyerseller; deal
Pooled Rewards & Incentives *(Model 3, 9)* reward; incentive; incentivize; contributor; contribution; curator; steem; curation; earn; pool; ecosystem; contribute; engagement; incentivizing; participant; stakeholder; encourage; earned; participation; valuable
Project *(Model 3, 61)* project; service; system; development; platform; use; market; number; token; stage; participant; payment; work; main; fund; create; company; receive; amount; account
Selected Detail-Level Topics and Content†
Peer-to-Peer Network (Model 2, 423) network; decentralized; centralized; peer-to-peer; participant; distributed; connect; incentivize; resource; incentive; built; access; run; infrastructure; running; allow; require; providing; build; without; secure
Other Associated Detail-Level Topics
Crowdfunding Campaign; Socioeconomic System; Ecosystem Development; Project Development; Platform Affordances; User Affordances; Foundation Governance; Rewards and Incentives; Passive Income; Fee Pool; Staking Mechanism; Loyalty Program; Reserve Pool; Voting, Decision-making; Ecommerce Marketplace; Provider Marketplace; Marketplace Transaction; Mobile App, App Store; Community

† Topic content presents the 20 words with the highest probability according to the FREX algorithm. The words are taken from the model indicated, and appear consistently across at least two of the three Mid-Level models.

tokens have made this sort of distributed capital raising possible, and have also served as a means of both distributing various kinds of participation (though not necessarily ownership) rights as well as building a set of potential users for each project with a vested interest in the project's success through their ownership of these coins.

The ability to generate project-specific tokens and cryptocurrencies is central to one of the primary mechanisms that projects use to coordinate distributed actors in the absence of formal contracts or organizational ties. The **Pooled Rewards and Incentives** topic describes the forms of social action that sustain distributed projects (e.g. "contribution," "curation," "engagement", "participation") as well as the mechanisms for incentivizing those actions through the granting of rewards. These rewards are typically worked out in game-theoretic terms by developers and then encoded into protocols that are themselves both a form and object of governance. I explore this mode of governance further in Chapter 7.

4.6 Discussion

The three technological clusters I've identified as general-purpose components of blockchain technology – computational infrastructure, programmable money and tokenization – capture the aspects of digital transformation that are distinctly enabled by blockchain technology relative to other technologies such as AI. The fourth cluster captures components of sociotechnical organization whose integration will require more analysis to cohere into a recognizable whole; I set that analysis aside for now and take it up again in Chapters 6 and 7.

The three technological dimensions in this chapter provide some initial insights into the nature of blockchain-based digital transformation. While I separate them here for analytical purposes (and do the same in Chapter 3's taxonomy for classification purposes), in practice the three are closely linked and often operate in combination when used in actual implementations. These joint uses reflect the clustered nature of general purpose technologies in general. More broadly, they articulate several core functions for the technology (providing generative infrastructure, creating and managing programmable money for payments, and creating and managing digital tokens representing value) that in turn offer a means of tracing variation in the ways those functions are used and combined across the wider ecology of implementations. This analysis is the core of the next chapter.

Chapter 5

The Role of Institutions in the Diffusion of a General Purpose Technology

Having established the components of blockchain as a general purpose technology in the preceding chapter, I now return to the question animating this dissertation: how does such a technology take its early steps into the world? In this chapter, I explore this question at the macrosocial level by examining how blockchain technology manifests across categories in the taxonomy that concludes Chapter 3.

I start from Rogers' definition of diffusion as requiring re-invention, meaning that new technologies are repurposed within each new setting in which they are adopted (Rogers 1995). In Rogers' terms, this is a "heterophilous" process marked by divergences in the uses and construction of the technology across settings.[1] Each of these will have their own requirements and modes of standardization, so that we would reasonably expect each to vary in what they use to reinvent their own version of the technology, and the outcomes will also vary in the extent of their use. One would thus expect that reinvention of blockchain technology through its development for various settings will draw on different aspects of the technology and will result in varying forms of digital transformation.

[1] *"One of the most distinctive problems in the diffusion of innovations is that the participants are quite heterophilous."* Rogers 1995, p. 19, emphasis in original.

It is a commonplace of Schumpeterian innovation that the "new" is, when dissected, an evolutionary recombination of what already exists.[2] Taking a strictly historicized view of innovation, however, risks eliminating the potential for novelty in technological components and thus limits the scope of possible change. Accounting for recombinant innovation with *new*, and especially new general-purpose, technologies thus requires an approach that allows for both new and existing components as the basis for recombination (Lester and Piore 2004; Stark 1996, 2009).

I address this by approaching the diffusion of a general purpose technology as the entry of a new set of components for recombination with each other and with what already exists, expanding the field of possible combinations. Approaching recombination in this way opens the door to questions that require further analysis, especially in the case of general purpose technologies that arrive in clusters. If the diffusing technology is being recombined, which components are being recombined, and with what? Rogers' work points to the need to attend to specific settings and requirements and to develop a way of attaching some kind of variation to the phenomenon. What technologies are being recombined? With what are they being recombined, and to what end? How do these vary across settings?

I frame this variation in what follows as different forms and degrees of *imbrication* of new technology with existing (and sometimes centuries-old) institutions, structures and processes. Imbrication takes into account that the diffusion of technologies doesn't happen as a simple binary of disruptive entrepreneurs versus incumbents, but rather of decisions about how a new technology might add more to an existing set of value offerings rather than simply reducing transaction costs (Frolov 2021). In order to give shape to my analysis, I follow Leonardi (2011) by defining imbrication as a specific combination of elements, which opens the door to an analysis of how such imbrication varies. Doing this allows me to work across categories to draw out broader patterns of imbrication across the detailed taxonomy I present in Appendix A. It also provides a structure within which I analyze the applications of the general purpose mechanisms

[2] For a biological view, see Jacob (1982, p. 35), who describes evolution as a process in which "new structures are elaborated out of preexisting components."

identified in the prior chapter and their contributions to the diffusion of blockchain technology and, by extension, of digital transformation.

I make two moves to bring out the answers to this question. First, I gather top-level categories from detailed taxonomy into clusters based on patterned commonalities in the ways their member projects combine the possible of new technology and the actual of existing practices and structures. I identify three such groupings, the first of which I term "digital economies" given that they each encompass existing industries and markets from the creation of raw materials through to the creation of end markets where consumers and producers meet. The second group – "digital finance" – spans the various finance-based applications of blockchain technology, including those focused on managing and trading assets, as well as those more closely focused on money. I call the final set "extra-institutional sources of trust" in deference to the pioneering work of Zucker (1986) on the relationship between large-scale social change and the transformation of the nature of trust. In contrast to the now well-established institutions Zucker studied, these new categories draw on computational and other digital methods to create variants on the same trust-producing institutions while operating largely outside of them.

Understanding the modes of change, and of imbrication, requires attention to the mundane infrastructures that are as much a zone of transformation as any app. Within each of the three broad clusters, we need to pay particular attention to the way blockchain technology interacts with existing infrastructures. Doing this infrastructural inversion is the only way to grasp the extent of both infrastructural and product-based digital transformation. Put another way, attention to infrastructure is necessary to encompass the full scope of generality of purpose of the technology.

In doing this, I identify three general patterns of reinvention. For Digital Economies, I find that reinvention is *pragmatic* or problem-oriented rather than fully transformational, and is focused on addressing existing problems that pre-dated blockchain and may ultimately be addressed through a combination of blockchain and other new technologies. By contrast, I find that Digital Finance is the site of *competitive* reinvention, in which both incumbent institutions and organizations and newcomers are working in parallel on fundamental changes to the global financial system. Finally, I find that projects in the Extra-Institutional Sources of Trust cluster are engaged in *speculative* reinvention in which the alternatives to existing institutions are being explored most intensively outside of those institutions.

5.1 Digital Economies and Pragmatist Reinvention

I use the umbrella descriptor "economies" for the largest grouping of top-level categories. These groupings have many of the hallmarks of sectors or industries in their incorporation of all stages of production, from sourcing basic inputs through to the construction of two-sided markets for consumers and producers of finished goods. I instead call them "economies" because of their added capacity to create their own programmable money. The viability of these economies is grounded in the affordances of blockchain technology, particularly in the creation of tokenized rights and marketplaces on which those rights can be traded (Catalini and Gans 2016). They also draw heavily on the technology's emphasis on validation and tracking to create authoritative records of provenance at various levels of granularity.

The dominant mode of reinvention for these projects is one of using the technology to address existing challenges in established domains, such as fraud in online advertising, rather than of wholesale transformation. I call this pattern *pragmatist* or problem-solving reinvention. This pattern arises in part because of the widespread impact of digital transformation also encompasses other technologies (e.g. virtual reality in online gaming, or payments in social media) and forms of sociotechnical organization (e.g. platform marketplaces in ridesharing) that are already well established. In this context, the generalized use of blockchain technology is for the creation of operational and business models in which participation is being changed to redistribute participation rights and rewards as a means of reorienting power and control away from a central actor or institution. These central actors can be monopsonistic firms in consolidating industries such as food or energy, or the more widely discussed platform monopolies in social media and the gig economy (Schor and Attwood-Charles 2017; Zuboff 2019). This is done by creating new forms of property rights, distributing those rights through incentives and transactions related to tokens that embody them, and creating markets and other platforms on which those tokenized rights can be earned and traded. Digital Economies projects marry this form of assetization (Birch and Muniesa 2020) through tokens with datafication that creates digital representations of both digital and physical objects. These digital representations in turn enable the tracking of provenance and authenticity not only of the end product (e.g. a song) but also of its components across their entire histories, allowing for more finely grained monetization of e.g. the percussion track of an individual song that is resampled in another.

The economies I describe in the detailed taxonomy at the end of Chapter 3 are clustered by the top-level classifications that comprise them. The first of these are *physical economies*, by which I mean economies in which the core mode of production involves physical objects and materials. The *physical production* classifications within this cluster include agriculture, energy and carbon and materials. In each of these cases, the role of blockchain technologies is primarily one of embedding data at an early stage of production into goods, and then tracking those goods as assets as they make their way into end markets, where they become tradable. *Agriculture*, for example, incorporates several projects involved in fishing whose primary activities involve creating collaborative supply chain networks for fisheries, canneries, retailers and other actors to establish the authenticity of fish based on their tracking through various supply chains. The *energy* classification is similarly focused on the tracking and trading of rights in renewable (primarily solar, but also hydro and wind) energy, though it also uses a combination of assetization and embedded sensors to allow participants to trade the energy they generate on local micro grids.

This cluster also includes classifications involved in the transportation of goods and people. Projects focused on the transportation of goods through shipping, freight, logistics and parcel delivery are in a sense extensions on the tracking and verification of goods in supply chains into their last stages of delivery to end clients. Because of this, these projects tend to involve the creation of their own blockchain ledgers. The lower data requirements for these projects, together with their focus on consumers as end users, means that they tend to operate as applications on existing blockchains.

The *creative economies* cluster comprises two top-level classifications linked by the common concept of creating and tokenizing definitive (and thus tradable) ownership rights in creative works. The models grouped *art and entertainment* into a top-level classification that split into separate topic-categories for *art and collectibles*, as well as *entertainment* (defined as music, film and television), each with their own concerns. The art and collectibles classification is dominated by concerns with establishing the provenance and authenticity of rare goods, as well as the creation and support of auction marketplaces in which tokenized forms are traded. Art projects also include a relatively new form of tokenization in which digital artworks are tokenized as unique or Non-Fungible Tokens (NFTs). Some projects go so far as to allow both the base layer of an image and its derivative additional layers to be tokenized (e.g. Async Art). The NFT creates uniqueness while also making it tradable. The *entertainment* category is more comprehensive in the activities in encompasses, which include everything from crowdfunding production and idea development to distribution by streaming and other methods. The core notion of property in these activities is the intellectual property rights of professional artists, and the projects in these classifications are centrally concerned with the creation of binding property rights as well as the ability for creators to track and monetize them effectively.

The *content and advertising* set of classifications is also primarily concerned with monetization, though the property involved and the market mechanisms differ. The *content* classification is centrally concerned with user-generated content (UGC), whether in the form of streaming video, social media, social networks, or other online forms. Monetization in these projects is typically through *advertising*, a classification that is so closely intertwined with content that the two are merged in smaller models. Advertising projects in the data set are almost entirely focused on online advertising, and in particular on using blockchain technologies to address both operational challenges (e.g. poor data, fraud) and economic problems (e.g. the oligopolistic power of platform incumbents Google and Facebook) challenges facing that industry.

The third cluster are the *gaming and virtual* set of economies. Reflecting the ascendance of online gaming (and longstanding role of internet gambling) as an economic force that now surpasses that of traditional entertainment in revenues, this cluster incorporates a broad scope of projects involved with the full spectrum of production, consumption, trading and other activities. This is most evident in the *gaming* category, which – like the existing industry – is split between platforms for game development and production of new games, marketplaces for game engines and components, platforms for gameplay, and exchanges for trading tokens, skins and other rewards earned in games. The *gambling* category is more focused on consumer-facing subcategories defined by online spaces for various kinds of gambling, betting, lotteries and jackpots. It also, however, includes projects that draw on various blockchain technologies to address core issues in online gambling. Key among these is the fairness of games. While, as the adage goes, "the house always wins" in Las Vegas and other professional gambling sites, blockchain-based gambling sites claim to draw on similar cryptographic technology as some consensus mechanisms for the generation of randomness (theoretically evening the odds for any individual of winning) that is both mathematically assured and verifiable.

The remaining two categories within this cluster touch on aspects of the prior two. Within the *sports* category, *eSports* and *fantasy sports* are among the fastest-growing segment of online gaming, which is reflected in their frequent grouping with *gaming* in models. This is particularly the case for *eSports*, where gameplay is often livestreamed so that the player can potentially earn revenue, followers, endorsements and other monetary rewards in the same way that gamers can (this is also the source of frequent overlap with *content* topics). This category also overlaps significantly with *gambling* via a heavy emphasis on *betting*, primarily on professional sporting events but also on eSports games. The *virtual worlds* category also overlaps with the others, and primarily *gaming*, given its focus on the development and use of various forms of mixed reality (Virtual Reality ("VR") and Augmented Reality ("AR")). The most widespread use of these technologies is in the creation of virtual gaming worlds, though the technology is also widely used in other categories such as *advertising* and *retail*.

5.2 Competitive Reinvention in Digital Finance

Digital Finance – the grouping of top-level Asset and Money Finance topics and sub-topics – is the part of the ecology of projects where fundamental transformations to the core institutions of the global financial system are being pursued simultaneously by incumbent organizations and challengers. These are parallel paths of innovation occurring in both peer-to-peer finance and existing state and private institutions, using variants of the same technology in an attempt to create their own versions of a digitized and tokenized financial system. Consequently, I term this pattern competitive reinvention.

At first glance, the categories in this grouping align well with the frequently used term "fintech," a growing set of digital innovations that are creating new products, services and business models in all aspects of finance (Philippon 2016). In keeping with this label, the subcategories within the top-level **Asset Finance** category include investment, trading and other means of capitalizing on the ability to create and trade digital assets. The bulk of the projects in this domain in my data set are involved in investment and trading of cryptocurrencies and cryptographic tokens. An increasing number have focused on tokenizing real estate and even stocks and bonds to make them tradable on token (rather than regulated equity and bond) markets and registered on a blockchain. At the same time, state and other actors have begun to issue new tokenized securities such as government debt in blockchain-based trials, bypassing traditional markets entirely.[3] While the latter remain in the experimental phase, they point to the potential of competitive reinvention to change the functioning of markets.

The top-level **Money Finance** category incorporates Cryptocurrency and crypto payments, as well as subcategories more traditionally oriented to other qualities of money beyond its use as a means of exchange. These are notably distinct from Cryptocurrency as a mid-level topic in the models, and appear in topics regarding monetary policy, gold and central banking infrastructure.

[3] For information on Thailand's retail bond offering in partnership with IBM, see https://newsroom.ibm.com/2020-10-05-Bank-of-Thailand-Launches-Worlds-First-Government-Savings-Bond-on-IBM-Blockchain-Technology. For information on the European Investment Bank's blockchain-based debt in partnership with Banque de France, see https://www.eib.org/en/press/all/2021-141-european-investment-bank-eib-issues-its-first-ever-digital-bond-on-a-public-blockchain.

Excavating the relationships between these topics requires attention to the nature of state currency and its pivotal role in the current financial infrastructure. Central banks are national entities with various mandates, but all are involved in providing central functions backed by their ability to issue and provide transactions in currency that they issue. Often called "central bank money" (and distinguished from "private" or "bank money" issued by banks based on deposits), this money is regarded as that country's risk-free asset. Larger central banks also manage national systems for the settlement of inter-bank transactions for payments and various kinds of securities transactions. The construction of such technology-enabled Real-Time Global Settlement (or RTGS) systems has been a focus for central banks around the world for decades.

The importance of settlement is easy to overlook precisely because of the success of these and other, private payment and settlement systems, such as Visa's. The ability to make transactions final, and to ensure timely payment from one party to another, is foundational to the viability of both money and of property rights, and thus of a functioning banking and investment system. RTGS systems play a special role among settlement systems in that they allow network participants to settle their transactions in central bank money, which means that the central bank (as the sponsor of that money) provides an added layer of risk protection for those able to access its systems.[4] This combination of state or central bank money and security has resulted in enormous amounts of money flowing through these systems. The Federal Reserve's RTGS system is called FedWire, and settled more than US$69 trillion of transactions in the month of January 2021.[5] The UK's system settled roughly half a trillion pounds of transactions daily in 2016, half of which were securities transactions.[6]

It is on precisely this terrain of central banks – the ability to produce "safe" money and to provide transaction finality – that cryptocurrency and blockchain attempt to intervene. In particular, so-called stablecoins, or cryptocurrencies seek to pair the peer-to-peer character and

[4] Smaller banks and payment networks often lack access to these systems (often called payment rails) and have to pay those who do to settle their transactions.

[5] https://web.archive.org/web/20210227131959/https://frbservices.org/resources/financial-services/wires/volume-value-stats/monthly-stats.html

[6] https://www.bankofengland.co.uk/-/media/boe/files/speech/2016/building-the-market-infrastructure-of-tomorrow-crest-rtgs-and-t pdf

payment mechanisms of cryptocurrencies with the relative price stability of state currencies, gold or money market funds. Stablecoins vary primarily in the means they use in their attempt to deliver price stability. The earliest variants defined "stability" in terms of the price of a more traditional asset – typically the U.S. Dollar or gold – and sought to peg the coin's value to that asset by using the proceeds of coin sales to purchase the same asset to ensure that the coin's price moved that of the asset. Another approach to stability involves using an underlying portfolio of cryptocurrencies which are actively traded, echoing the approach of traditional money market funds.

Beginning in 2020, a new wave of stablecoins and other peer-to-peer finance offerings came to market under the banner of "DeFi," or Decentralized Finance. Elements of DeFi as a peer-to-peer form of finance existed prior to the popularization of the term in 2020 (particularly in P2P lending and insurance), but it wasn't until that year that it began to cohere in practice and in identification as a coherent set of practices and business models. The core impetus of DeFi is the replacement of centralized actors in financial markets and systems with solutions and practices that claim to be more decentralized. The centralized building blocks DeFi seeks to replace are generally hidden from nonprofessionals but include infrastructural providers of liquidity and capital, as well as credit scoring, trading clearinghouses and other forms of financial infrastructure. In DeFi, these are typically replaced with some combination of algorithmic (e.g. automated rather than centralized market making) and collective alternatives. The collective alternatives vary in their goals but typically provide incentives for individuals to bond or stake their tokens, and then aggregate those staked holdings into larger pools that provide liquidity for the trading mechanisms and exchanges that are at the core of the offering.

Like other DeFi projects, these new forms of stablecoins are largely based on some combination of algorithmic management and sophisticated arbitrage trading, which the projects rely on for executing whatever monetary policy they enact in the pursuit of price stability. In many cases, these involve creating an underlying portfolio of collateral and then offering multiple tokens with different rights in that portfolio (e.g. an equity token with capital appreciation rights and a bond token with income rights), distinct from the stablecoin itself. The tokens are actively traded on a project-specific exchange allowing both long and short positions (rather than being traded against a variety of other tokens, as occurs on most exchanges), and their prices are designed to play a significant role in valuing the underlying portfolio and thus (ideally) supporting the value of the stablecoin linked to the portfolio. This combination makes DeFi stablecoins relatively complex entities, with aspects of bank accounts (interest-bearing deposits denominated in a stable currency) and speculative investments (actively traded sub-tokens with debt and equity characteristics).

The claim of stablecoins to provide a price-stable cryptocurrency operating outside of the traditional monetary system has made them a focus of central banks that are already in the midst of significant changes to payment and settlement systems. Central bank and related institutions, particularly the Bank for International Settlements (BIS), often called the central bank of central banks, have responded to the rise of DeFi and of DeFi stablecoins with a rapidly growing body of research and pilot projects on central bank digital currencies or "CBDCs," digital variants of central bank money (Auer and Böhme 2020; of England 2020; Chaum, Grothoff, and Moser 2021).[7] As a rationale, most of this research cites advances in payment networks and the advent of stablecoins, particularly the widely followed efforts of Facebook's Diem (originally Libra) project, along with the declining use of physical cash. Many of these projects also link blockchain-based advances in central bank digital cash to blockchain-based efforts to innovate on existing RTGS systems for central bank settlement, though this is not universal.

Two of the most advanced pilot projects capture some of the range of possibilities in this research. In Sweden, the central bank has been among the most active in the world in exploring new cash and payment systems given the country's unusually low use of physical cash (Riksbank 2021). This shrinking use of physical cash threatens to reduce the Riksbank's role in the economy since that cash is currently the only form of central bank-issued money available to individuals.

[7] For a list of countries engaged in various aspects of creating CBDCs from the BIS see https://web.archive.org/web/20210420131520/https://www.bis.org/publ/work880.htm. For a list from IMF researcher John Kiff, see https://web.archive.org/web/20210420131723/https://kiffmeister.blogspot.com/2019/12/countries-where-retail-cbdc-is-being.html. Both sites are as of April 2021.

To address this, the Riksbank has explored an alternative payments network in which individuals can access tokens linked to the value of the Swedish Krone but held in individual accounts, and available to be used in payment networks. The companies involved in these networks would be the nodes in the blockchain-based network and have access to Sweden's RTGS system "RIX," where they would be able to conduct settlement in e-Krone. Given its focus on use by individuals, the e-Krone project is considered a potential retail CBDC.

At the other end of the use spectrum are wholesale CBDCs, whose users would be financial institutions focused solely on settlement. Given the global scope of many of the banks operating in these settlement networks and the relative size of their markets, wholesale projects often contemplate the possibility of both domestic as well as cross-border (and thus multi-currency) settlement. One of the most advanced projects in wholesale, cross-border CBDC grew out of initial pilot projects for blockchain-based settlement platforms initiated by the central banks of Singapore and Thailand. In 2021, these projects merged into a larger, collaborative effort spearheaded by the Bank for International Settlements exploring a new concept they call m-CDBC, or multiple CBDC. The core concept of the project is the use of digital state currencies (available only to financial institutions with access to central bank settlement) to make cross-border settlements more efficient. As with the e-Krone project, the nodes in this closed network would be run by approved financial institutions rather than open to individuals.

5.3 Creating Extra-Institutional Trust through Speculative Reinvention

Of all the institutional characteristics invoked in discussions of blockchain technology, *trust* is perhaps the most cited. While some scholars and practitioners describe the technology as "trustless," others write more convincingly about the production of new kinds of trust. The apparent contradiction resolves with further consideration of the relational structure these authors assume. Those claiming trustlessness focus on the ability of A to transact with B without either party having to trust the other to complete the transaction. This dyadic interpretation of trust appears compelling at first glance but doesn't hold up to scrutiny in its implication that blockchain is unique among technologies in not requiring what Giddens called "system-level trust" (Giddens 1990), or trust in a complex system whose inner workings aren't knowable to the non-expert.

A more satisfying approach to trust claims that blockchain technology shifts the locus of trust from existing intermediaries to a distributed set of across actors in a blockchain network (Seidel 2018). This claim is similar in spirit to Zucker's (1986) pioneering work on the production of new modes of trust during another period of transformation, namely the industrialization of the United States. These new modes, which Zucker called institutional-based trust, were made necessary by the disruption of traditionally localized market interactions based on interpersonal trust. The new modes replaced interpersonal trust with structures that formalized rules protecting the interests of both exchange parties and operated independently of any single exchange. This independence in turn allows them to scale in ways that support increasingly complex exchange across social and geographic boundaries. Key among these are the creation of various intermediaries (e.g. banks, insurance companies, financial markets) and intermediary functions (e.g. escrow, formal contracts) that allow strangers to transact via what Shapiro (1987) calls "guardians of impersonal trust."

Unlike Zucker, I don't have the benefit of a century's distance to say with assurance that blockchain technology is institutionalized in a way that can replicate institution-based trust. What is apparent in my data set is its potential to do so through a number of mechanisms that replicate trust-producing functions of existing institutions beyond the level of individual transactions. These are sources of *extra-institutional trust*, a term I apply to the final categories of the taxonomy in Chapter 3. The projects providing these alternatives to traditional sources of trust ground their claims of effectiveness in the ability of blockchain ledgers to provide tamper-free records, and in the automated execution of increasingly complex smart contracts, though each combines these in a different way. Because these projects seek to provide novel alternatives that envisions a future of working outside of traditional institutions, I define them as involved in *speculative reinvention*.

The first top-level category in this grouping collects projects that involve the establishment and verification of truth claims, many of which are traditionally provided by Zucker's institutions such as notaries and state records. I term this grouping **Identity and Verification**. The identity-focused projects seek to fill a gap in the Internet's structure and functioning, neither of which originally incorporated an identity layer. The solution of many projects is to build on some combination of attestations to an individual's identity claims from traditional state records (e.g. a passport or driver's license) and from individuals and organizations taken as authoritative within that project's framework. These attestations are then used as the basis for a verified form of identification that can theoretically be used to address regulatory concerns around KYC (Know Your Client) rules for new customers seeking entry to blockchain platforms, as well as for individuals to avoid the inconvenience of having to verify their identity more than once. The blockchain being used provides an authoritative record of this identity, while its cryptographic security ensures the record's uniqueness and security.

These same tools are applied to the broader question of establishing some form of "truth" as well as providing participants with the means to verify this truth by projects involved in verification. As with identity, most of the projects in this subcategory are focused on either building on existing institutions (such as notaries) who provide the initial attestation that then is digitized and verified within the system, or replacing them entirely by creating purely digital forms of verification. In both cases, these projects rely heavily on not only the security but also the uniqueness of cryptographic hashes.

One of the primary motors of institutional trust in Zucker's study is created at the individual and firm level by membership in a recognized profession or subgroup with a set of internally agreed standards. These standards provide signals of legitimacy to both members and non-members (Starr 1982), while also providing actors a means of demonstrating their adherence to them through credentialing. A generalized version of this mechanism is at the root of the next Top-Level category, which encompasses several subcategories concerned with **Claims and Validation**. The most prominent subcategories in this set are Education and Employment.

Although the former includes projects focused on more traditional forms of education, the bulk of projects in it are focused on the development of specific skills, and on establishing trustworthy attestations of those skills through official records/transcripts. The Employment subcategory mirrors this emphasis on measurable and verified skills but from the perspective of demand for those skills, especially in the context of the task orientation of gig employment (Cedefop 2020). The platform or two-sided marketplace form of sociotechnical organization is especially pronounced at the intersection of these two subcategories, where such marketplaces provide a mechanism for matching worker skills with employer task requirements.

Zucker (1986) theorized that the institutionalization of the various trusted forms of exchange she identified would lead to the development of markets in which those forms were traded. Perhaps the most prominent of such forms in the context of the Internet is reputation, which has long been studied as a basis for trust in other settings (Greif 1989). In particular, ratings and reviews have become a central means of establishing reputation, and for consumers or buyers seeking to navigate markets or platforms in which their information is limited (Diekmann et al. 2014; Kuwabara 2015; Resnick and Zeckhauser 2002; Schor and Attwood-Charles 2017). Within the ecology of blockchain implementations, projects focused on building reputation systems are seeking to create reputations that are secure and verified, but also portable from one setting to the next. This latter characteristic would allow individuals to incorporate their amassed reputation from all of their activities online into a portable form of identity, a possibility not currently feasible given the forms of enclosure central to the business models of platform incumbents (Zuboff 2019).

The final top-level domain in this cluster is **Governance**. Like the other classifications I group under "Extra Institutional Trust," this domain includes projects that provide mechanisms that are also in use in one form or another in other domains (e.g. all protocols necessarily incorporate aspects of governance). These projects are distinguished by their formalization of these mechanisms and their focus on providing them to other projects as product offerings.

The first of these are focused on the use of market and computational methods to navigate uncertainty and include both prediction markets and platforms for dispute resolution. Although blockchains are based on the secure storage of data, they ironically have no native mechanism for accessing data in the real world beyond what is entered directly into the ledger. Such data is particularly important for smart contracts, which are self-implementing code linked to blockchains whose execution is often conditional on some event in world (e.g. "pay Alice $100 if the temperature exceeds 75 degrees"). Such contracts rely on a construct called an oracle that feeds data to the smart contract, typically from an authoritative external source. The linkage between oracles and smart contracts echoes those based on APIs or application programming interfaces, which standardize data linkages between actors on the web.

The potential for reliance on centralized data providers, as well as the general market orientation of cryptoeconomics, has made prediction markets a common solution to this challenge. Although forms vary, prediction markets typically combine some form of exchange on which predictions are traded and an oracle to establish the definitive outcome. The trading of predictions or bets on the market establishes a market price. Some prediction market protocols (such as Augur, a market on the Ethereum blockchain) also incorporate a decentralized oracle based on other market mechanisms. In all cases, the assumption that participants are incentivized to converge on the correct outcome typically means that predictors are rewarded for alignment with the majority.

Another method for navigating uncertainty is through the development of methods for resolving disputes that can't be satisfied within the constraints of code, particularly in the domain of contractors and gig employment. Projects focused on *Dispute Resolution* draw on a combination of digital methods, reputation and verified expertise to provide adjudication of such disputes. While the digital mechanisms are largely new, the institutions they build on include juries, law and expert arbitration.

The final subcategory in this group is a locus of active experimentation at the intersection of all of the GPT aspects of the technology I describe in the prior chapter. Distributed Autonomous Organizations, or "DAOs," are sociotechnical forms of organization that seek to retain the benefits of meso-level organization while replacing most of the substance of Weberian bureaucracy with code. This substitution is enabled by the increasing sophistication of the smart contracts that comprise this code, as well as by the visibility of ongoing experimentation with the form, which has created an unofficial knowledge base regarding its implementation.

As with other areas in blockchain, DAOs are often built on enabling platforms developed specifically for the creation and launch of DAOs. In addition to ongoing protocol and other technological development, these platforms can also provide other services for dispute resolution that range from digital jurisdictions to the creation of juries and "courts" in which participation combines tokenized incentives with smart contract-defined decision processes. Perhaps the most widely known of these is Aragon, an application and protocol running on the Ethereum blockchain that offers tools for the creation of DAOs (Aragon Core) as well as a jurisdiction (the Aragon Network) for collective dispute resolution through courts and juries who are incentivized by both monetary and reputation rewards. Aragon hosted 1,755 DAOs on its network by April 2021.[8]

5.4 Discussion: The Role of Technological-Institutional Alignment

The forms of reinvention I describe in this chapter point to two broad findings regarding the nature of interactions among general purpose technologies, institutions, and digital transformation. The first is the qualitative differences in the relationship between GPTs and institutions in each of the three patterns. The affordances of programmable money, tokenization and computational infrastructure (as well as its ability to provide finality) are definitionally broadly salient given my data. What makes them especially salient in their imbrication is appears to be related to the extent of overlap between the nature of the general purpose technologies and the institutional functions for which they might be used.

[8]https://web.archive.org/web/20210422141756/https://scout.cool/aragon/mainnet, visited and archived on April 22, 2021.

From an institutional perspective, the most dramatic potential transformations I describe above are the result of the closest alignment between the general purpose aspects of the technology and the functions of institutions involved in the transformation of a domain such as banking or law. As a first approximation, the greater this alignment, the more likely the pattern of transformation will result from either competitive or speculative reinvention, which is to say, the more likely the pattern will appear to involve a fundamental revision to existing institutional arrangements. Why some institutions respond with their own innovations (in competitive reinvention) while others don't (leading to speculative reinvention by those outside) is a question for further study. By contrast, domains in which the uses of the technology are largely for solving well-established problems of coordination in existing industries are less intensely engaged with underlying institutions.

The other dimension along which these patterns vary is the extent to which other digital GPTs are involved in digital transformation. Here again, technological-institutional alignment may play a role, though seeing it clearly requires expanding the perspective to include other digital GPTs. The various incumbent industries I describe above as economies would appear to complicate the role of blockchain as a general purpose technology by limiting its role in transformation in these settings given the more central role being played by the Internet of Things, machine learning and other digital GPTs. A closer look, however, reveals that blockchain technology is not sufficient in these settings to be transformational on its own, but can contribute to larger processes of transformation being driven by other technologies. Those other technologies are likely more closely aligned with institutionalized functions (e.g. the Internet of Things aligning with global supply chains) than is blockchain technology.

These patterns of reinvention and imbrication at the macro level are an important step toward answering the question of how a new GPT travels in its early stages, but the analysis thus far begs further questions that can only be answered by a closer look at the meso-level mechanisms involved in these implementations. How do projects organize to implement this new technology? Once organized, how are they governed as ongoing entities? The following chapters address each of these questions, respectively.

Chapter 6

The Microstructure of Blockchain Implementations

The preceding chapters have approached blockchain technology as a potential GPT from the macro level. Doing so has required bracketing much of the “organizational and institutional arrangements” (Bresnahan and Trajtenberg 1995, p. 101) through which these technologies move from their initial development into broader development and use. While scholars have begun to approach technologies of digital transformation from this perspective (Goldfarb, Taska, and Teodoridis 2020), the more generalized question of how these arrangements work in practice remains unanswered. I use this chapter to contribute a blockchain-based answer to the question of how a new GPT is implemented in its early, pre-settlement stages.

Answering this seemingly simple question requires a resolution of ambiguities surrounding both the technology and the forms of organization involved in implementing it. The term “blockchain” has been used so widely that it has become, as one practitioner described it, “a semantic wasteland” (Carter 2018). The situation is little better in academia. Scholars often use “blockchain” either without reference to the other necessary technologies that comprise the technology or as though Bitcoin is the exemplar. The latter is especially misleading now that blockchain has expanded so widely. Economic studies of ICOs and other investment issues do better at accommodating some of the heterogeneity of applications, though they also tend to flatten the technology to investment time series or even use “ICO” as an organizational unit rather than a capital-raising event. This approach also eliminates the wide range of private implementations from consideration.

Fixing the unit of analysis thus requires accounting for the ”fractal-like” nature (Carlaw and Lipsey 2002) of the technology given the constituent components I articulated in Chapter 2. I do this by focusing on implementations of the technology rather than on any single component of it. Doing so fixes my analysis at what Leonardi (2009) terms “the implementation line” separating the earliest development and later widespread adoption of a settled technology, a locus that necessitates a sociomaterial analysis that proceeds from the assumption that the technology and the social organization involved in implementing it are integrally linked.

This approach also implicitly addresses the other unit of analysis problem, this one posed by the technology’s embedding in current practices in technology development. Technology development is normatively open source in blockchain, making open source communities obvious organizational structures and use of the term “community” omnipresent. At the same time, framing technology as a platform embedded in a larger relational ecosystem (an organizational counterpart to individuals in an open source community) has also become a normative conceptual framework for technology development in general, and thus for those developing blockchain implementations. As a result, there is also a widespread practice of referring to implementations as platforms, open source protocols or communities, even though these are often individual pieces of a larger sociomaterial whole, and thus are inadequate to the question I ask.

The prevalence of these terms is obvious in my data set and in the analysis of the prior chapters, particularly in the fourth aspect in Chapter 4 that I termed Sociotechnical Organization. I reproduce part of Table 4.5 below for reference.

This conceptual haze is mirrored by an ambiguity in the organizational literature, which is increasingly rich in studies of new decentralized structures for organizing that are part and parcel of 21st century technology development and implementation. In addition to the traditional

Table 6.1: General Purpose Aspects of Sociotechnical Organization in Blockchain

Mid-Level Topics & Content†
Crowdfunding *(Model 2, 67)* crowdfunding; backer; project; ito; campaign; startup; fundraising; teammember; funding; milestone; entrepreneur; translation; raise; projectteam; english; tokenholder; team; completion; financing
Ecosystem *(Model 1, 26)* ecosystem; adoption; economy; foundation; vision; innovation; incentivize; sustainable; mass; drive; engagement; leverage; leveraging; accelerate; enabling; engage; community; align; adopter; utility
Marketplace *(Model 2, 60)* seller; buyer; marketplace; vendor; auction; bid; supplier; listing; escrow; invoice; goods; mediator; item; ebay; bidder; buying; sell; goodsandservice; buyerseller; deal
Pooled Rewards & Incentives *(Model 3, 9)* reward; incentive; incentivize; contributor; contribution; curator; steem; curation; earn; pool; ecosystem; contribute; engagement; incentivizing; participant; stakeholder; encourage; earned; participation; valuable
Project *(Model 3, 61)* project; service; system; development; platform; use; market; number; token; stage; participant; payment; work; main; fund; create; company; receive; amount; account

† Topic content presents the 20 words with the highest probability according to the FREX algorithm. The words are taken from the model indicated, and appear consistently across at least two of the three Mid-Level models.

organization, these include open source code and user communities, project teams, platform structures, ecosystems and other relatively recent innovations (Obstfeld 2017; Gawer 2014; Jacobides, Cennamo, and Gawer 2018; O'Mahony 2002; O'Mahony and Lakhani 2011; Parker, Van Alstyne, and Jiang 2017).

Although each of these building blocks have been studied extensively individually as "the basic organizational unit" (Grabher 2004, p. 2) for each of these various types, we know less about organizational design decisions by entrepreneurs for whom these function as a *set* of possible building blocks to be incorporated combinatorially (Grandori and Furnari 2008, 2013). This matters because these component forms (e.g. communities, platforms) rarely operate in isolation but are typically part of larger assemblages that raise the much larger question of how coordination works across these components. For example, what is Linux? If Linux is distinct from its community, it is unclear in much of the literature whether that is because "Linux" is a conceptual container for multiple organizational components that include the community, a non-profit foundation, and the Linux codebase, or whether each of these is a separate organizational entity.

This sort of agglomeration is not unique to open source software. Gulati et al. (2012) ask us to:

> ...consider Apple, which has reduced its own direct internal inputs into some of its devices, while launching an immensely successful Apps Store, virtually none of whose products it produces, but 30 percent of whose revenues it captures. Apple has also moved the kernel of its core operating system to an open community even as it partners with Intel for aspects of its hardware.... (Gulati, Puranam, and Tushman 2012, p. 572)

Like the Linux example above, this description of Apple paints an intriguing picture that decenters the firm Apple Inc. by making it only one of several organizational structures that include the App Store market platform, the ecosystem of app developers and users, an open source community and an increasingly disaggregated supply chain. Like Linux, it isn't entirely clear what organization is identified by the name "Apple" anymore, and by extension, who is doing the "launching" and "building" in the quote above.

The paradox here is perhaps clearest in the paths organizational scholars have taken to theorizing an organizational construct that accounts for their ambiguity. Among the most relevant is the meta-organization, which Gulati et al. define as "networks of firms or individuals not bound by authority based on employment relationships, but characterized by a system-level goal" (2012, p. 573). Scholars of communities, platforms, ecosystems and other forms have drawn on this construct in attempting to define the forms they study, reflecting the construct's origins in attempting to account for their general characteristics (O'Mahony and Lakhani 2011; Gawer 2014; Jacobides, Cennamo, and Gawer 2018). At the same time, the quote above was drawn from the same paper, in which the authors also apply the term to Apple. This fractal aspect of meta-organizations composed of other meta-organizations creates new challenges beyond those inherent in the component structures. Staying at the implementation line, especially for a technology that is itself unsettled and in its relatively early stage of development, also means attending to heterogeneity at multiple levels: in the technological and organizational components chosen for a given implementation, in their joint configurations, and over time as the technology and lessons from earlier experiments continue to evolve. The nature of membership and of boundaries become more complex and layered, as traditional employment-based membership for individuals in a single organization is expanded to included membership in more diffuse communities and ecosystems, even as these collectives themselves negotiate membership in the larger meta-organization. While these collectives include formal organizations, these formal organizations are neither fully in control, nor are they necessarily the central actors. This structure recasts one of the fundamental problems of organization: how does such an organization maintain coherence when it is meta all the way down?

To answer this question, I draw on recent work on organizational design from the microstructural perspective, which proceeds from the assertion that any form of organizing can be characterized by its sub-components or microstructures, as well as the way these microstructures are combined into a goal-oriented whole (Puranam 2018; Puranam, Alexy, and Reitzig 2014). I begin by identifying generalized modes of coordination and integration that have been identified in the literature on various forms of meta-organizations as a way of gathering potential mechanisms in the following section. The subsequent section turns to blockchain technology and builds up the set of microstructures available to organizational and technology designers. Section 6.3 combines the insights of the prior sections to develop a framework I used to analyze a series of implementations of the technology. Section 6.4 concludes with a generalized framework for analyzing the organization of blockchain implementations.

6.1 Coordination and Integration in Meta-Organizations

Attempting to identify organizational arrangements for the implementation of a 21st GPT requires a different toolkit than the more traditional Weberian approach of building a single ideal type. The success of 20th century theorizing in organizational sociology about established organizational forms operating in fields and ecologies is unlikely to be as useful for a technology in its early stages, let alone for the "menagerie" (Sabel and Zeitlin 1997) of organizational structures now in use. The question of how to identify meso-level organization has become further complicated in the digital era given that the relationship between organizational boundaries and resources are under constant renegotiation (Neff and Stark 2004), and can even be "inverted" by the wholesale shifting of key resources outside of the organization, and also outside of the reach of traditional means of securing them (Parker, Van Alstyne, and Jiang 2017; Watkins and Stark 2018).

The fact that these features have been linked to organizational structures that have been broadly grouped under the banner of meta-organization indicates at least some form of commonality, though that commonality may not be apparent given the siloed literature on open source and other communities, platforms, ecosystems, alliances, consortia and other forms. Shifting from an analytical search for established forms to meta-structure would thus appear to be a swap of one problem for another, namely to trade institutionalization for an unsettled morass of heterogeneity (Glynn, Barr, and Dacin 2000; Grandori and Furnari 2013). However, as the examples of Linux and Apple show, the components of sociotechnical microstructures have themselves been well studied, and it is on that basis that I proceed by identifying three mechanisms of coordination in these structures.

6.1.1 Mediating Hierarchies

Both the Linux and Apple examples show that formal organizations (e.g. the Linux Foundation or Apple Inc.) don't disappear but rather remain as important coordinating actors within these larger meta-organizations. Even the most resolutely anti-hierarchical communities eventually evolve some form of organizing (as distinct from a definitive organizational form) that enables their functioning (Burton and Obel 2013; Dobusch and Schoeneborn 2015; O'Mahony 2007; Massa and O'Mahony 2021). Formal organizations can be penumbral (Massa and O'Mahony 2021; O'Mahony and Lakhani 2011) or central actors (Dhanaraj and Parkhe 2006), or even develop randomly (Burton and Obel 2013), but nonetheless retain an important role in stabilizing the larger whole.

Whatever their structure, the active role of these formal organizations within the meta-organizations they enable is fundamentally one of coordinating ongoing resources from external stakeholders. While there are deep literatures on venture capital, IPOs and other forms of capital raising for new firms, and on securing production resources on external markets, the shift to meta-organizations necessitates several shifts in resource mobilization and coordination. First, it implies a temporal shift in emphasis from initial capital-raising to an ongoing process of mobilizing and orchestrating resources that aren't traded on external markets but also aren't subject to strict controls through the employment relation or formal contracting. Second, it fixes the locus of these resources in a nebulous zone that may be outside of the formal bounds of the organization but not fully separated into an external environment. Finally, and perhaps most important, the examples of Linux and Apple show that the external communities of stakeholders are likely divided into distinct groups with potentially conflicting interests. As a result, the primary question of coordination in meta-meta-organizations is one of mediating the interests of not only one but many types of actors, necessitating attention to the specific interests or "regimes of value" of the stakeholders whose participation is necessary (Barrett, Oborn, and Orlikowski 2016; see also Brandtner 2017).

Several streams of literature have examined the role of formal organizations in mediating between multiple communities, and thus multiple groups of stakeholders (Barrett, Oborn, and Orlikowski 2016; Blair and Stout 2005; Blair and Stout 1999; Klein et al. 2019). A relatively new branch of corporate governance literature, based in part on Elinor Ostrom's work, approaches the question of corporate governance through the lens of resource coordination across multiple stakeholder communities (Blair and Stout 2005; Blair and Stout 1999; Klein et al. 2019). In this literature, the challenge for firms is not simply securing resources, but rather creating unique resources for the organization in partnership with stakeholders. These unique, firm-specific resources are created through a process of co-investment and co-creation between the firm and those stakeholders willing to commit their own resources (in whatever form) to the focal organization's use. The focal organization's challenge then becomes one of presenting both stakeholder-specific incentives and rewards, as well as balancing the interests across all stakeholders and the firm itself, in order to accomplish core objectives.[1]

This complex role of mediating between multiple stakeholders leads these scholars to call them "mediating hierarchies." I adopt this term in what follows to refer to any formal organization that operates as a component of a larger meta- (or even meta-meta) organization, and focus my review of the literature on the mechanisms these mediating hierarchies enable. Put another way, I examine the non-hierarchical coordination of resources by (or at least involving) formal hierarchies.

6.1.2 Resource Mobilization

The ongoing mobilization of external resources can be seen as a challenge of transforming a public with similar interests to a community working toward a shared goal (O'Mahony and Lakhani 2011; Starr 2021). More organization-centric work posits a single firm or organization around which communities organize. They may organize as communities of practice or learning for individuals (Brown and Duguid 1991, Lave and Wenger 1991), or as organization-centric communities of problem-solving and innovation (Boudreau and Lakhani 2015; Lakhani, Lifshitz-Assaf, and Tushman 2013). In the latter case, the central organization (e.g. NASA) draws on the expertise of an external community to stimulate innovative solutions to seemingly intractable challenges.

The rise of crowdfunding and crowdsourcing of resources has become the focus of a growing body of research. Examples include traditionally entrepreneurial funding for startups as well as building new technology and new products, artistic production, social causes, scientific research, and other endeavors (Agrawal, Catalini, and Goldfarb 2014; Hui, Greenberg, and Gerber 2014; Mollick 2014; Wheat et al. 2013). One consistent finding in this work is that the problem of bootstrapping funds for some quantum of work is transformed into a larger challenge of bootstrapping and maintaining a community. Success involves cultivating a community through

[1]This concept is typically presented in the language of strategy, in terms of the economic value proposition for stakeholders, how (and how much) economic value is created through the venture, and how the economic value created is divided between parties.

articulating a compelling vision for the project, and frequent communication on progress toward that vision (Agrawal, Catalini, and Goldfarb 2011; Mollick 2014).

Whether they engage or crowdfunding or other means, one way for emergent organizations to mobilize resources is through visions of the future, expressed in terms of the imagined affordances of a technology (Beckert 2016; Nagy and Neff 2015) as well as the modes of organization involved in implementing it. The latter are typically expressed in the form of *models* (Doganova and Eyquem-Renault 2009; Teece 2010; Zott, Amit, and Massa 2011). These narratives do a particular kind of work by expressing a theory about the world and a hypothesis about how best to act on that theory (Nickerson and Zenger 2004; Felin and Zenger 2017). In doing so, they serve as a means of coordinating disparate actors through the articulation of models around which actors can choose to coordinate (Strang and Meyer 1993).

6.1.3 Network Orchestration

In such settings, formal organizations tend to assume a “network-weaving” (Ingram and Torfason 2010) role involving orchestration rather than direct control (Dhanaraj and Parkhe 2006; Giudici, Reinmoeller, and Ravasi 2017). One of the primary modes of orchestration in open source communities is the creation of “architectures of participation” (Baldwin and Clark 2006) that define processes for contributing to and reviewing code. While the earliest work on open source took a micro perspective to ask why skilled developers would participate in uncompensated work (Lerner and Tirole 2001, 2002, 2005), more recent work has focused on the meso level and on the increasing formalization of organizational processes that shape participation (Eghbal 2020; Rozas and Huckle 2020; Baldwin and Clark 2006; O’Mahony and Ferraro 2007). These processes typically include a detailed set of instructions on the parameters of a viable proposal, as well as the review process.

Perhaps the central contribution of such architectures is their articulation of role structures, which help stabilize interactions in non-hierarchical settings in which tasks are interdependent and formal structures are relatively weak (Weick 1993). Role structures are particularly important in creative industries such as film production, as well as in the ongoing viability of community-based productions such as Wikipedia (Arazy, Lifshitz□Assaf, and Balila 2019; Bechky 2006). In spite of the mythos that anyone can contribute, every open source community of scale has developed a form of hierarchy that controls the path of decision-making and generally constrains votes on code changes or additions to a formally defined and typically small subset of the community, and the ability to make those changes to an even smaller subset of maintainers (MacCormack, Rusnak, and Baldwin 2006; Shaw and Hill 2014).

The basic definition of platform and its ecosystem specifies a network based on three roles: the provider of the platform, a set of "complementors" on the platform, and a still larger set of end users of both those offerings and the platform itself. (Adner 2017; Wareham, Fox, and Cano Giner 2014) The canonical example is Apple's App Store, for which Apple provides a technological and market platform on which app developers build, and within which end users make consumer decisions about the apps. In this case, the "ecosystem" consists of the platform (App Store) and either the apps themselves (if looking from a technology perspective) or the makers of the apps (if looking from an economic or market perspective). From either perspective, the platform and the apps/app makers are viewed as complementary components of a larger whole that only has economic value for end-users if both layers are present, though end users are not considered part of the ecosystem itself.

The effectiveness of mediating hierarchies (typically, the owners of large market platforms) in orchestrating resources is conditioned on their ability to balance the divergent interests of complementors and end users. One of the primary means of doing this is formalizing the incentives, rewards and sanctions into economic and technological models (Eaton et al. 2015; Ghazawneh and Henfridsson 2013). Another approach is through the use of reputation, scoring and ranking mechanisms that discipline participants (Fourcade and Healy 2017). Although companies such as Uber, Facebook and Airbnb tend to use the language of community when describing their users, in practice these platforms mediate multi-sided markets or ecosystems in which members are economic actors rather than community members as the term is commonly understood. These platforms seek to build on widely studied online coordination mechanisms such as reputations, scoring and ranking to determine the value of contributions and provide rewards and sanctions (Resnick and Varian 1997).

Mediating hierarchies in open source software also act as network orchestrators by looking beyond the coordination of their codebase to the larger ecosystems of organizations in which they operate. Historical studies of open source foundations and have tended to emphasize their buffering role in shielding coders from distraction by managing contractual relations with external parties (O'Mahony and Bechky 2008, O'Mahony and Ferraro 2007). Since those studies, the institutionalization of open source has led to increasing resource demands in terms of functionality, even as it has shifted the emphasis of foundations from an inward focus on coders to an outward need to coordinate among the organizations that use the codebase and (ideally) contribute programming time to its ongoing development. One of the primary means of accomplishing this is through active grant and other funding programs for external organizations and developer teams, who often compete for funding to develop specific functionalities.

The final form of network orchestration is the creation of a new mediating hierarchy by a group of incumbent organizations. These external hierarchies are typically created to solve coordination problems involving multiple firms that are either unable or unwilling to devote the resources necessary to accomplish a particular task. At the smaller end of the scale, the project has become a standard format for working across functional boundaries in both traditional and creative industries (Grabher 2002; 2004Obstfeld 2017; Sydow and Staber 2002; Sydow, Lindkvist, and DeFillippi 2004) Although these projects operate as independent entities, they are embedded in larger interorganizational networks and epistemic communities that provide material resources and knowledge necessary for the project to meet its goals.

In contrast to these often temporary forms of meta-organization, incumbent organizations in technology and other industries often cooperate to create consortia, alliances and other interorganizational forms in which the incumbent organizations themselves are the members (Ahrne and Brunsson 2011). These forms are particularly common in cases where industry actors coordinate around the creation of new protocols and standards (Rosenkopf and Tushman 1998).

6.2 The Microstructures of Blockchain Technology

"Blockchain," as I have suggested throughout this dissertation, is neither a single technology nor a settled term, but instead has come to signify a set of technologies that have emerged out of distributed systems design, cryptography and game theory.[2] As I noted in Chapter 2, distributed systems are typically defined as 1) a set of autonomous "nodes" defined by a combination of hardware and software, each of which 2) contributes to a single "process" that 3) appears to its end users as a single system or interface.

At its most general, the term "blockchain" itself refers to a ledger of time-stamped transactional data that is distributed across a network in which no one node (or node owner) holds sole control over the data. This definition begs several questions that reveal the need to think of "blockchain technology" in systems terms rather than as a whole: Who uses the network, and to what end? Who is allowed to run a node in the network? How do the nodes reach agreement? Who pays for the network?

These questions are remarkably similar to those needing to be solved in traditional organization. Distributed systems face the same organizational problems as meta-organization in general, and in fact heighten them - information sharing, incentives, and assessing control tradeoffs are fundamental to dispersed systems of autonomous agents who must somehow coordinate to accomplish a goal, whether those actors are humans or boxes in a server farm (Fox 1981; Puranam 2018). This similarity means that implementations of blockchain technology lend themselves particularly well to a sociomaterial approach in which both the technology and its social organization are treated as integrally related.

In what follows, I use this approach to build up the basic components or microstructures that appear in blockchain implementations. I describe these components as parts of an organizational

[2]For a comprehensive technical overview of the core technologies, see Narayanan et al. 2016).

vocabulary or toolkit rather than universally used. I begin with the resources that are typically involved in implementing the technology, along with the groups of stakeholders who co-create those resources, then describe the variety of mediating hierarchies involved in this process, and end with the core of projects.

6.2.1 Defining Resources and Stakeholders

The most basic microstructures fall along two dimensions that roughly track Thompson's (1967) technological and institutional dimensions of organization. The first of these is a sort of technological axis that links *code* and *computation*. The classic definition of distributed systems in computer science is a set of networked nodes that are simultaneously engaged in the same process while maintaining their autonomy (van Steen and Tanenbaum 2017; Lamport 1978). Although this definition implies a concretized "node" as a physical point in a network, the computer science literature defines nodes as hybrids constructed of both software and the hardware on which it runs.

Because of this, two obvious resource requirements are code and hardware that jointly comprise the nodes in the network. The fact that blockchain software is so new, and its functions so complex, means that software engineers, architects and developers are central - and often the dominant - actors in blockchain projects. Much of this complexity stems from the nature of distributed systems, which require that the code that runs on the nodes in the P2P networks generally includes algorithmic mechanisms that enable them to create authoritative records (or ledgers,) in the absence of institutionalized – and often hierarchical – mechanisms enabling transactions (Zucker 1986).

Blockchain developers have a wide range of skill levels and employment arrangements but are generally part of the larger open source software development community. Open source has been central to blockchain since the development of Bitcoin, which was itself an outgrowth of cryptographic, cypherpunk and other communities enmeshed with open source code development (Brunton 2019). Github has since become institutionalized in the blockchain ecology as the de facto site for code development and storage, as have the open source practices and structures described in the prior section. Key among these are the formalized processes for proposing, deliberating and deciding on changes to the underlying protocol. Python established its own version, the Python Enhancement Proposal or "PEP" process, in 2000; the practice was later adopted by BitTorrent (BEPs, beginning in 2008), Debian (DEPS, in 2009) and other open source projects.[3] Bitcoin and Ethereum went on to establish BIPs and EIPs processes, or Bitcoin Improvement Proposals and Ethereum Improvement Proposals, respectively.

As an inherently network-based set of technologies, blockchains and the applications that build on them rely on various forms of hardware to provide the necessary data inputs, perform the computation that defines the network's function and enable end users to participate. This makes the owners of the hardware on which the computations are run particularly important stakeholders.

Given the influence of Bitcoin as a prototype in the popular press and imagination, mining nodes are the most widely known form of hardware in blockchain. By providing the computations used to secure blockchain ledgers, these computers collectively provide "hashpower," or the large-scale computational output that would theoretically be cost-prohibitive for any one actor to attempt to dominate. While this approach remains central for Bitcoin and several other projects, it is far from the only use of hardware in the overall ecology. The data gathering and transmission capabilities of distributed devices are often the basis for the extension of blockchain projects into the physical world through supply chains in agriculture, energy, logistics and other domains. Some of the latter draw on distributed networks of remote sensors (e.g. RFID tags in supply chains, solar panels and sensors in energy grids, bluetooth chips in Internet of Things devices), mesh networks of internet routers and mobile phones, drones, distributed computation (e.g. cloud services, distributed storage, file sharing) and other technologies. More complex projects incorporating these devices and their output require attention to more diversified networks of devices, and the needs and interests of those who own them.

[3] For Python, see https://www.python.org/dev/peps/; BitTorrent http://www.bittorrent.org/beps/bep_0000.html; Debian https://dep-team.pages.debian.net/; Apache Kafka https://cwiki.apache.org/confluence/display/KAFKA/Kafka+Improvement+Proposals

Often subsumed under discussions of nodes, hardware and open source, the resources and actors captured by Thompson's (1967) institutional axis are no less important to the actual implementation of the technology. The first of these are providers of capital and other kinds of funding for both the development and ongoing operation of these complex projects. One predominant form of capital raising for *de novo* projects in my data set is through Initial Coin Offerings (ICOs), a form of crowdfunding in which startups sell tokens to the public in return for capital to fund the project's buildout. Because the same tokens are often used for raising funds and for working within the project, ICOs theoretically and empirically solve two closely connected problems for technological projects based on P2P networks. The first of these is of course funding, while the second is bootstrapping a community of hardware owners and other stakeholders willing to provision the needs of a new network.

Venture capital firms are also heavily involved in funding blockchain startups, and take a range of approaches. In some cases, they participate in a project as ecosystem-oriented community members, buying tokens (albeit often at a discount prior to the token sale), participating in mining and other distributed tasks, and occasionally funding development grants for other projects or teams building complementary technologies. In other cases, these investors will purchase equity in a for-profit firm set up in part as a mediating hierarchy to accept those investments.

This sort of ecosystem funding behavior is also characteristic of larger projects in the space, which often reverse the typical entrepreneurial flow of capital into organizations, and instead fund external developments that are beneficial to both their own business and to the larger ecology. These are typically undertaken through grant programs that target key resources that can include gaps in a project's technological functions, or the funding of open source developers who can focus on a given project full-time as a result of funding. As a result, the funders of aspects of one project (particularly its code development) are often another project in the same ecosystem, which typically means those working in the context of a single blockchain platform.

The final institutional category are users, though this category is complicated by the institutional nature of the technology, and of its designers' conceptions of it. Rather than single "products" (such as a cryptocurrency), most blockchain projects describe their offerings as platforms, though their language doesn't always use this term with a consistent meaning. For purely infrastructural projects, the platform is largely technological and involves the creation of a framework on which others can develop their own projects. In such cases, "users" are more aptly termed ecosystem partners or (in the language of strategy) "complementors."

In other cases, the platform is a multi-sided market. As Catalini and Gans (2016) note, one of the primary contributions of blockchain technology is a reduction in the costs of forming new markets by effectively eliminating verification costs for actors in transactions. Market offerings in my data set are diverse in their application across domains but tend to share the multi-sided market structure (Catalini and Gans 2016; Lerner and Tirole 2002), with buyers/consumers, sellers/producers, reviewers and moderators making contributions that are compensated with tokens. As a result, the set of "users" differs significantly from the traditional users of a given technology and looks more like what sharing economy companies term "partners" in that users are also generally expected to participate in the market in some way, with compensation via tokens for their contributions. This reflects the nature of 21st century economic platforms, which include a number of roles beyond those of the archetypical buyer and seller. The difference between blockchain and prior iterations of the web economy (e.g. "Web 2.0" or "the participatory web") is that compensation is often extended to even the smallest interactions through microtransactions enabled by smart contracts.

6.2.2 Mediating Hierarchies in Blockchain

As with Apple and Linux, the overall ecology of blockchain projects is awash in formal organizational entities that act as coordinating (rather than governing) actors within these meta-organizations, with foundations and for-profit corporations being the most frequent forms. These hierarchies play a central role in coordinating the networked communities and ecosystems of actors (organizational and individual) that - together with the project core and the hierarchies themselves - comprise the meta-organization.

Following the precedent set by large open source projects, and later adopted by Bitcoin and Ethereum, blockchain and infrastructural-level projects have tended to establish non-profit foundations that often echo key functions of open source foundations such as holding any patents, licenses, investment capital and copyrights in the name of the project (O'Mahony and Bechky 2008).[4] Blockchain project foundations are typically funded out of the proceeds of ICOs, making them among the largest holders of a given token in a project's early years. Foundations often use a portion of these funds to compensate Core Devs for their ongoing development of protocols and code.

Foundations play a number of roles in coordinating the development and testing of code in tandem with the larger technology community using practices and formal structures developed by the open source community. As in the larger open source field, these processes are typically (though not always) managed by the foundations of projects. One of the primary means of doing this is in the early days of an implementation through the creation of a beta versions of the project's network, typically called a *testnet*. A testnet provides an opportunity for a project's Core Devs as well as external developers to experiment with and debug new features, as well as an opportunity for hardware providers and external coders to contribute to the process, all with much lower stakes than are present in the full network.

The foundations of larger and better funded projects are also actively engaged with the larger technology community, typically in ways that seek to foster larger ecosystems of developers building on and extending their respective protocols. This form of mobilization extends the development of code beyond the focal project to include larger ecosystems of developers and projects. This happens most frequently in the form of periodic grant competitions for external projects and teams of developers competing to build functional extensions to the underlying protocol. Mediating hierarchies typically coordinate these activities by establishing and

[4]The name is more important than it might seem given the propensity of open source blockchain projects to "fork" into separate variations. When these forks are contested, ownership of the original name provides a powerful counter-argument to claims by those initiating a competitor to be operating in the true spirit of the original, and thus to be the "real" version.

communicating technical priorities and soliciting proposals based on those priorities.

Another form that appears in projects is the registered corporation, which often co-occurs with foundations. In their minimal form, these corporate forms are purely instrumental legal constructs that provide a vehicle for the bureaucratic necessities of accepting outside venture capital, managing payroll, contracts, etc. Blockchain project-related corporations are also often a means for the Core Devs to monetize their skills by developing tailored applications for established corporate entities interested in the project's technology. Another variant involves corporations building infrastructural projects that expand the functionality of (and access to) the blockchain project on which they focus. Such projects are often developed or heavily sponsored by private firms such as Blockstream (for Bitcoin) or ConsenSys (for Ethereum).

Where foundations and corporations are the most prevalent forms involved in the launch of new projects, established/incumbent actors typically work through consortia or other forms of non-profit associations (such as Facebook's Diem project and its Libra Association). These forms are most prevalent in settings in which established formal organizations are committed to participating in the development and support of a project but lack either the resources or the risk appetite to undertake it alone. This form came to prominence in technology in the 1990s as a form of risk sharing in research and development (Evan and Olk 1990) and remains a preferred structure for enabling industry-wide collaboration over the development of technological and regulatory standards.

6.2.3 The Project Core

I define the core of a blockchain project as comprising the core team as well as the economic and technological models that express the project's goals and the expected path to executing them. Core Teams are involved in the creation and design of projects and their specifications and tend to grow in size and diversity of functions as projects grow. At the outset of new implementations, the core team includes the "nascent entrepreneurs" (Ruef 2000) who articulate the earliest versions of a project's stated goals and the methods for achieving them. The core team are typically some (often overlapping) combination of co-founders, developers, marketers, community managers, and others, depending on the nature of the project. Perhaps the most pervasive Core Team members are core developers or “Core Devs,” who are typically the programmers, cryptographers, economists and others who develop the initial technical specification, economic incentives and protocol of the project.

The other components of the project core are the technical and economic models that define the project. A project’s *technical model* is typically embodied by its technical specification, which is a description of the architectural design and functionality of the project’s technology. Project protocols are in practice assemblages of other components that include base protocols, algorithms, cryptographic primitives and other components. The individual recombinations are specific to each project, but the components that are recombined typically come from peer-based commons in science and code (Benkler 2013).

Economic models in blockchain projects typically shift the emphasis from value appropriation by a central economic actor to one of distributed value creation as well as mechanisms for more closely linking value capture to contributions. This shift represents a broader imperative in platform economies and ecosystems to realign models in a way that makes both possible while retaining the viability of the hub firm (Adner 2017; Adner and Kapoor 2010).

The economic and technical models function as statements of goals and identity by establishing the project’s objectives and the means toward achieving those objectives. This function is why I place them at the core of projects.

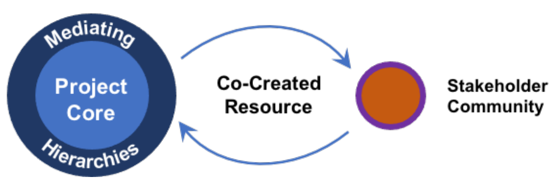
6.3 Empirical Analysis: Microstructures and Configurational Examples

My empirical approach for the remainder of this chapter draws on multiple case studies as a means of developing a more generalized insight into the sociotechnical configurations involved in implementing blockchain (Eisenhardt and Graebner 2007; Yin 2009). I ground this analysis in the insights of the prior two sections by first creating a generalized microstructure, and then articulating configurational variations on that microstructure in the empirical cases that follow (Grandori and Furnari 2008, 2013). I then work across those cases analogically to abstract a general configuration (Vaughan 2014).

6.3.1 Generalized Microstructure

I begin my empirical analysis by establishing a generic microstructure based on my review of the three mechanisms of coordination in the literature on meta-organizations, as well as the microstructures of coordination in blockchain. Figure 6.1 positions the Core of each project in relation to external communities of stakeholders. The arrows between each capture flows of information, capital and other resources between those communities and the project core. These flows are bidirectional, reflecting the co-creation of the core resources described in text inside the arrows.

Figure 6.1: Blockchain Project Microstructures



This generic microstructure captures the broad outlines of the prior sections, and provides a basis for articulating and comparing the specific configurations in the cases that follow.

6.3.2 Empirical Cases

My approach to case selection is heavily informed by the work in prior chapters in identifying the nature and breadth of implementations at the macro level. Rather than working within the taxonomy I developed in those chapters, however, I instead work with another concept derived from distributed systems design: layering. Layering is a vertical metaphor for the partitioning of systems into functional layers that carry out distinct aspects of the system's work, with each layer supporting the functions of those layers above it. From an organizational perspective, the concept of layering is implicit in the nature of networked ecosystems in their combination of a platform layer and a layer of applications. It is also inherent in the nature of blockchain technology as an enabling digital technology (Iansiti and Lakhani 2017; Teece 2018).

I combine concepts to consider blockchain technology as being implemented in three general layers, each of which involves the development of a distributed system with increasing complexity as more layers (and thus more stakeholders) are added.

The Blockchain Layer

The first of these is the *blockchain layer*, which is what most discussions of "blockchain" implicitly refer to, and captures the infrastructural aspect of blockchain I described in Chapter 4. The blockchain layer begins with a peer-to-peer (P2P) network of computational nodes engaged in the process of generating an authoritative chronological record. This most often means a record of ownership, identity, provenance or some other claim regarding a digital (or digitized, in the case of physical objects) asset.

The primary form of variation at the level of the blockchain layer is based on who can run peer nodes in the network. Given that the code is open sourced, this typically reduces to a question of who is allowed to run a server that participates in the P2P network, and thus has access to the ledger. Networks that are closed to only approved members are called *permissioned* networks, while those that are mandated as open to anyone willing and able to run a node are known as *permissionless* networks. As a first approximation, permissioned networks are the primary structure used by corporate, state and other regulated actors, while permissionless networks are used by cryptocurrencies and the other structures I describe below.

In cryptocurrencies, node networks typically follow the Bitcoin example of being permissionless. Given the potential for hostile or "byzantine" actors to join and direct the network toward their own ends, the chief concern tends to be the security of the overall network. Permissionless networks typically address this concern through a combination of rewards and (occasionally) penalties that are encoded into their consensus algorithms and processes.

Permissioned networks are theoretically simpler to mobilize and control since they control who can join, though they bring a different set of implementation and coordination challenges. The primary one is not technical but rather political, namely, how to share sensitive information with competitors, regulators and other outside parties without compromising security. For enterprises, these problems are often addressed through participation in consortia, a long-established mode of coordination for industry peers and their partner organizations to mobilize jointly to solve collective challenges.

Example: MediLedger

In the pharmaceutical industry, a recent change in the regulatory framework has triggered the creation of a consortium of industry and related firms who joined to develop a blockchain project known as MediLedger. The effort was spurred by the U.S. Federal Food and Drug Administration's 2013 decision to require tracking of pharmaceuticals across their entire lifecycle by the year 2023.[5] Doing so will require the cooperation of not only drug manufacturers but also pharmacies, logistics companies and others. The MediLedger project was formed in 2017 to explore the feasibility of such a cooperative effort led by industry and is structured as a consortium that includes representatives from most major drug manufacturers, pharmaceutical benefit managers, national retailers, and both Fed Ex and UPS.

As articulated by its members, the objective of MediLedger is to meet the FDA's objectives while incorporating the need for privacy of the consortium's members. Operationally, they accomplish these aims by protecting the record of transactions on a permissioned blockchain using a combination of cryptographic methods that allows for complete tracking as a package moves from one participant to another, without revealing the contents to the entire consortium (maintaining strict privacy of business data) or involving a centralized agency (maintaining decentralization).

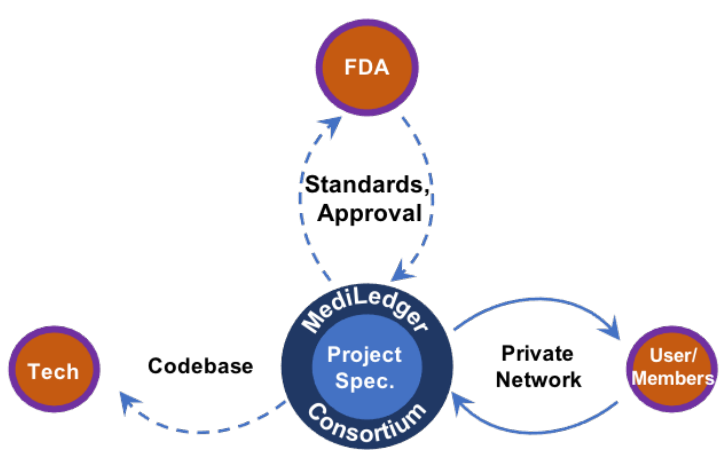
[5] For more information, see https://www.fda.gov/drugs/drug-supply-chain-integrity/drug-supply-chain-security-act-dscsa.

Accomplishing these joint goals is possible because of the network's dual structure in which each member organization runs its own Private Node for network communication, and has the capacity to also run a Consensus Node (defined by separate software) that participates in confirming new transactions and adding them to the blockchain. The Consensus nodes use smart contracts to verify that each serial number (for each transaction) can be linked back to verification of its authenticity to its original manufacturer, as well as having a complete chain of ownership from that manufacturer to its present state, while concealing other information using zero-knowledge proofs, a recent advance in cryptography.

This in turn requires establishing uniqueness. Each package is given a unique serial number by its manufacturer, which it retains across its lifecycle. Individual serial numbers are represented in the system as Non-Fungible Tokens, or NFTs, a structure designed to create a unique (and hence non-fungible) digital asset representing a unique physical or digital object. In this case, the unique object is the serial number for the package being tracked. Changes in ownership of the package in the physical world are thus represented as changes of ownership of its digital or NFT representation on the blockchain.

The structure of the MediLedger project appears in Figure 6.2, with the consortium as the sole mediating hierarchy and the project's technical and operational specification at its core. The bulk of the project is concentrated in its private network, for which its users are also the members providing the servers on which the nodes run.

Figure 6.2: MediLedger Project



A portion of the code developed by the technology firm that also helped coordinate the consortium (Chronicled) was posted to the firm's Github account, though it was developed offline and seems to have been posted there for storage than for ongoing development. As a result, I only include a dotted line leading from the consortium to the tech community, and do not include an arrow representing that community's interaction with the code.

The Blockchain Platform

The function of blockchains as enabling technological infrastructure became formally established in the modern sense of a platform with the launch of the Ethereum project. Ethereum was designed as an evolutionary step beyond Bitcoin's limitations. Where Bitcoin was designed to make computation expensive for the sake of security, Ethereum was designed to harness its network participants' aggregated computational power for the execution of *smart contracts*, applications of computer code with a misleading name in that they are not definitively legal contracts as their name would imply (Szabo 1997). Instead, they are a means of encoding the logic of increasingly complex "real-world" processes, automating their execution, and memorializing their results on a ledger, all through computer code. This linking of actions in the world to the automated authority of an underlying ledger is central to the digitization of those processes, and also to the technology's reach across so many domains.

This structure of a blockchain as an enabling, general-purpose technology platform supporting an increasing range of apps with their own tokens (in distinction to the older Bitcoin structure of supporting a single cryptocurrency) has become a dominant form, as platforms such as EOS, Qtum, Tezos and others have each emerged as platform competitors claiming to offer their own advances in technology, governance and other dimensions. That said, it is far from the only structure.

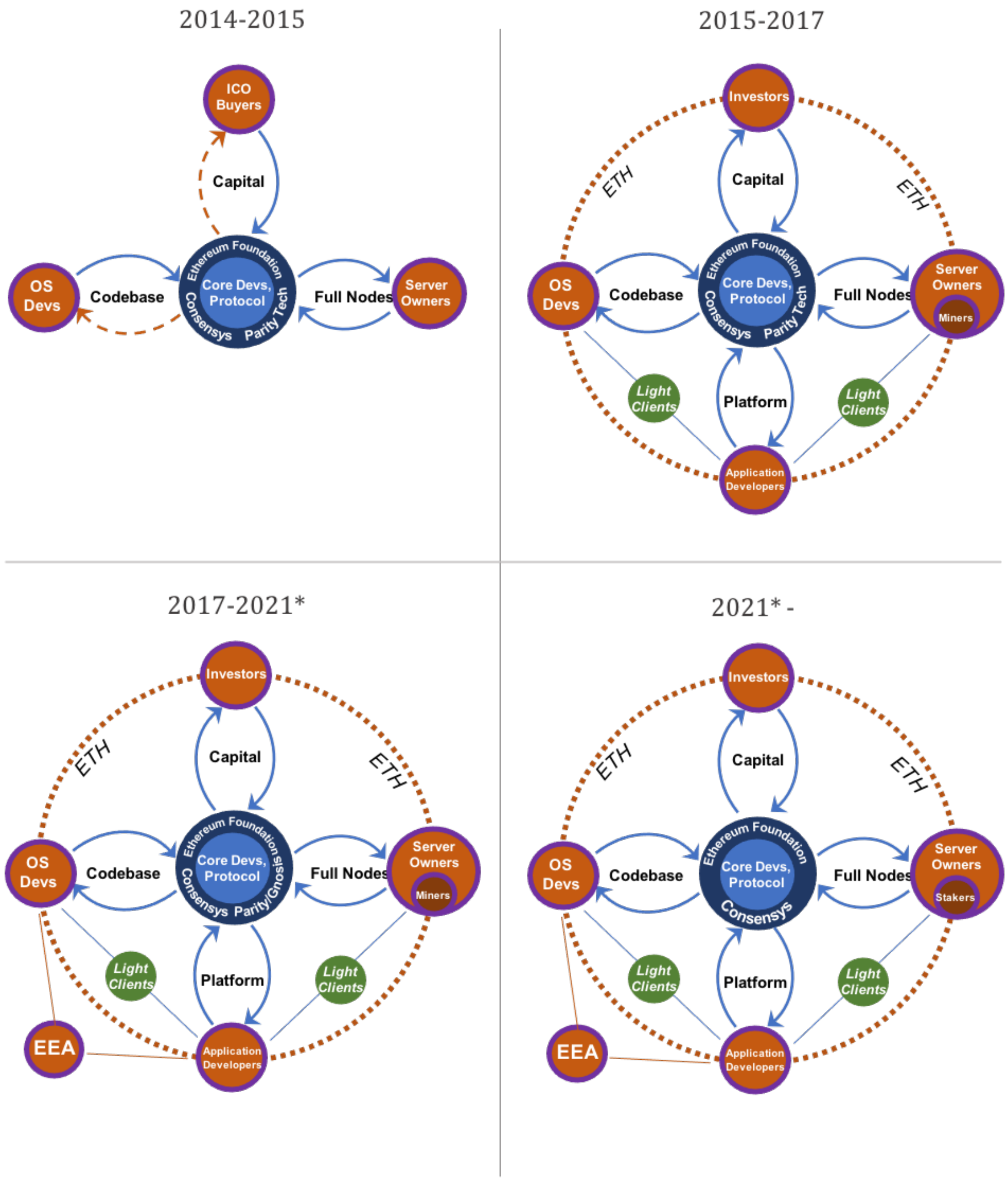
This structure and function are enabled by an expansion of the functionality of the economic and technical models, and further involvement by mediating hierarchies. Blockchain platforms often take this imperative further through their ability to create implementation-specific tokens or money, with the parameters of that money formalized in the project's protocol. This ability is used in a constitutional (Desan 2014) manner that is central to the notion of creating new and decentralized economies in which the project manages incentives through the use of encoded game-theoretic coordination mechanisms. This is accomplished in blockchain projects through a model of monetary management based on the number of tokens, monetary policy around money supply, and the attaching of token payments to incentives encoded in smart contracts (Cong, Li, and Wang 2019a; Cong, Li, and Wang 2019b).

Example: Ethereum

The Ethereum project is exemplary of the category of Blockchain Platforms. It was the first blockchain project designed specifically for this purpose, was one of the first primary ICOs, and also has consistently attracted among the largest communities of open source developers and new applications. The organizational configuration of the project is accordingly more complex than the others in this paper and has changed significantly over time. Figure 6.3 presents a stylized view of the project

Ethereum's Project Core has from its inception consisted of a team of Core Devs as well as a group of researchers working on ongoing improvements to the protocol, as well as the technical specification. The general protocol and technical specification were initially presented in two papers (the "white" and "yellow" papers, respectively) written by cofounders Vitalik Buterin (2013) and Gavin Wood (2014). The intellectual property of the Core is largely owned and buffered by the Ethereum Foundation, a Swiss non-profit funded from part of the proceeds of Ethereum's initial coin offering, making it one of the largest owners of ETH tokens. Two other organizations, each founded by a self-described co-founder of Ethereum, have also played significant roles in the project, though these roles have changed over time. The for-profit firm Consensys was founded by Joseph Lubin, one of the co-founders of Ethereum, and has been closely involved in the project since then. Parity Technologies was founded by Gavin Wood, another Ethereum cofounder and Ethereum's initial Chief Technology Officer, and played a

Figure 6.3: Ethereum Project Over Time



central role in Ethereum until 2020.

The upper left, and earliest, segment of Figure X shows that these three organizations took the role of mediating hierarchies in the first days of the project following its precedent-setting ICO in 2014. Purchasers in the ICO paid Bitcoin in return for receiving Ethereum's Ether or ETH tokens, which were also distributed to key actors in the ecosystem.[6] The Ethereum Foundation was funded from these proceeds, and has since been integrally involved with the ongoing development of the Ethereum protocol and code, and to that end, pays the salaries of many of the Core Devs as well as a dispersed group of researchers working on evolving the protocol. In the project's earliest days, it funded hackathons and "bug bounties" for developers to test and identify vulnerabilities in the project's code, which included code to be downloaded by those running nodes in the P2P network on which the ledger ran.

The upper-right panel of Figure 6.3 reflects the project's increasing complexity as it moved from its earliest testing stages to the transition to next version of the protocol (named 'Frontier") in 2015. This was effectively a fully functional version of the project as originally envisioned, and included several expansions. First, the pool of open source developers divided to include developers working on applications to run on the Ethereum blockchain (the bottom circle), in addition to the OS community working on Ethereum's own code. These developers were the primary users mentioned in Buterin's original white paper. As with any technology platform, these complementors contribute to the value of the overall Ethereum platform by building new applications using the Ethereum network. These applications are comprised largely of smart contracts, which in turn rely on the computational process of block production in the Ethereum ledger for execution.

The 2015 release also added a new group of actors within the overall set of providers of nodes. Following the Bitcoin example, Ethereum went live using proof-of-work (and the longest chain rule) as the core of its consensus algorithm. Those with specialized hardware capable of competing in proof-of-work mining, known as miners, thus became a subset of the server owners in the Ethereum network.

The incorporation of new groups of technological actors in Ethereum was accomplished through the creation of new *clients*, or software specifications that can be used to run a node. In Ethereum, clients are differentiated by type of user, and by the coding language used, reflecting

[6] This approach is known as pre-mining.

the broad range of actors in the larger ecosystem. These include, among others, nodes (and thus clients) tasked with storing the ledger's history, known as full nodes, and others that allow app developers and other interested parties to access the network without needing to store history. The latter are known as light clients, and necessary for developing applications that draw on the blockchain without participating directly in it.

The next stage of Ethereum's history (the lower left panel of Figure 6.3) was one of extraordinary expansion in the project's overall ecosystem due to a combination of an enormous speculative bubble in ICOs between 2017 and 2018 and several more durable developments. Key among the latter was the creation of a consortium of industry incumbent giants interested in developing their own, likely private, versions of Ethereum. The Ethereum Enterprise Alliance, or EEA, now one of the largest consortia in blockchain, has developed specifications for several private variants on Ethereum's technology to be used in permissioned networks. ConsenSys played a major role in building and coordinating the EEA, and (through a later acquisition of JP Morgan's Quorum group) of providing a client and reference implementation specifically for this use.

In late 2019, Parity Tech announced that it was withdrawing its developer resources from the reference Ethereum client they had developed and maintained since Ethereum's launch, the second most widely used. They also announced that they would be transitioning the client to a community orchestrated through a DAO, or decentralized autonomous organization. The resulting OpenEthereum project was ultimately overseen by another important project called Gnosis, using a significant grant from the Ethereum Foundation. Because of this, and because of the importance of the client, I include both Parity and Gnosis among the mediating hierarchies for that period.

The final, lower-right panel in Figure 6.3 has two subtle but fundamental differences from the third. The first is the consolidation of the mediating hierarchies to the Foundation and ConsenSys, reflecting the withdrawal (announced June 1, 2021) of Gnosis from supporting the former Parity client. The second change reflects a far more consequential shift, perhaps the most consequential protocol change in the overall ecology to date. Ethereum is preparing to change from proof-of-work to proof-of-stake sometime in late 2021 as the basis for its new approach to blockchain consensus. Where node operators have until now been paid in a similar manner to those in the Bitcoin network, for providing intentionally wasteful computation, the Proof-of-Stake approach would limit participation in block production to a much smaller group of actors who would be selected algorithmically based (in part) on the amount of Ethereum tokens they bonded or "staked" in order to participate. In addition to the enormously reduced environmental impact of this approach, it will also (due to various technological parameters being included) offer potentially better security.

The Blockchain-Enabled Platform

In practice, many of the most influential projects implementing blockchain technology are not themselves ledgers but instead draw on the computational infrastructure provided by established blockchain projects such as Ethereum or EOS. These projects are designed to capitalize on key features of blockchain protocols, while also extending their functionality. As a result, they are aligned with the infrastructural nature of blockchain technology (Iansiti and Lakhani 2017), but are not themselves blockchains. Those focused on more general solutions to technological problems such as blockchain's lack of scalability are often called "Layer 2" solutions in the field, though this is not a well-defined term. In other cases, projects that are domain-specific in their focus on creating generative platforms for third party developers in a specific domain.

Whatever their purpose, these platforms add a second P2P layer on top of the underlying ledger's P2P network. This layering creates interdependence that is enabled in large part by the connections between protocols at each level, which must be constructed in such a way that they offer tools and structures that are sufficiently specialized that they work on specific platforms, but general purpose enough that they allow for open innovation on the part of those building on the next higher level. This proliferation of interests has led many projects to introduce multiple tokens geared to different functions and actors, though even those with a single token are linked to the tokens of other projects on a given platform through their shared reliance on the same infrastructure.

Example: Golem

Golem defines itself as a computational marketplace where those seeking large-scale computational resources traditionally available from centralized cloud providers such as Amazon or Google can operate in a P2P environment. The platform uses Ethereum's infrastructure for payments between members but has its own P2P network of hardware owners who are paid to provide their unused processing capacity to those requesting the capacity to execute specific computational tasks such as rendering animation or testing a new cryptographic tool. Golem also offers its computational platform to developers of third-party applications seeking to capitalize on its resources. The project coordinates these parties by providing SDKs (software development kits) and other tools.

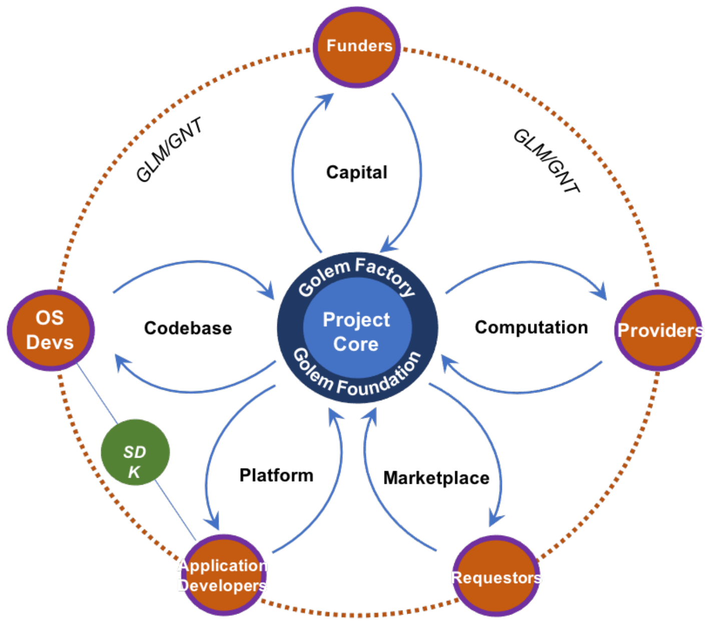
Figure 6.4 presents Golem's project structure. As for Ethereum and MediLedger, the horizontal axis consists of an open source developer and technology community interacting with Golem's evolving codebase, as well as a P2P network of servers and their owners who provide computational processing capacity. The user communities in the bottom circle incorporates the buyers of processing (called "requestors" by Golem) as well as application developers, who are supported by the provision of SDKs (software development kits) and other templated tools in via a technology platform.

The core of the project consists of a team of founders and core devs and other professionals, some paid by the for-profit entity Golem Factory and others working as advisors, as well as the project's protocol and technical specification. The latter have evolved significantly, most recently to shift Golem to a new kind of Ethereum-based payment system that is more efficient than relying on the main Ethereum blockchain. The other mediating hierarchy, the non-profit Golem Foundation, was founded as a distinct, for-profit entity in 2019 with the goal of developing more experimental mechanisms to support the project's long-term economic growth.[7]

---

[7]"This step is … the Golem Foundation, which will strive for new — perhaps innovative and experimental, and at the same time riskier — approaches to the value proposition for Golem and for the Golem Network Token

Figure 6.4: Golem Project



Payments and other means of coordination within the project are made using two project-specific tokens. The first of these, GLM, was issued in 2017 as part of the project's highly successful ICO in that year. The relatively early date of the ICO makes Golem one of the most established of the thousands of projects working on the Ethereum blockchain. It also, however, means that GLM was created before the Ethereum project formalized its ERC20 token standard. In addition to streamlining the process of creating a new token for projects using the Ethereum blockchain, the ERC20 standard also helped standardize fungibility between tokens, making ERC20 tokens much easier for investing, trading and other activities relying on moving funds from one token to another. To address this gap, the Golem project introduced the ERC20-standard GNT token, and at the time of this writing is in the process of replacing the GLM token with GNT. [8]

Golem has also played an indirect role in the orchestration of the larger Ethereum ecosystem. Like several other Ethereum-based projects, Golem secured significant funding during the ICO

(GNT). This includes testing new hypotheses and looking for attractive solutions that potentially increase the value of the entire project in the future." https://web.archive.org/web/20210623161554/https://golem.foundation/2019/06/28/introducing-golem-foundation.html

[8]For more information, see https://web.archive.org/web/20210623163216/https://blog.golemproject.net/gnt-to-glm-migration/.

boom, and began to use portions those funds in concert with the Ethereum Foundation in order to continue funding Ethereum's development. Golem did this in part by pooling its donations with those of other leading Ethereum infrastructural projects such as Gnosis to create the Ethereum Community Foundation.[9] Six of these organizations collectively donated more than $2 million in 2018 to create a pool of funds governed by the Ethereum Foundation. The ECF has since made grants to dozens of projects, teams and individual researchers building new technology.

6.4 Discussion: A Configurational Menu

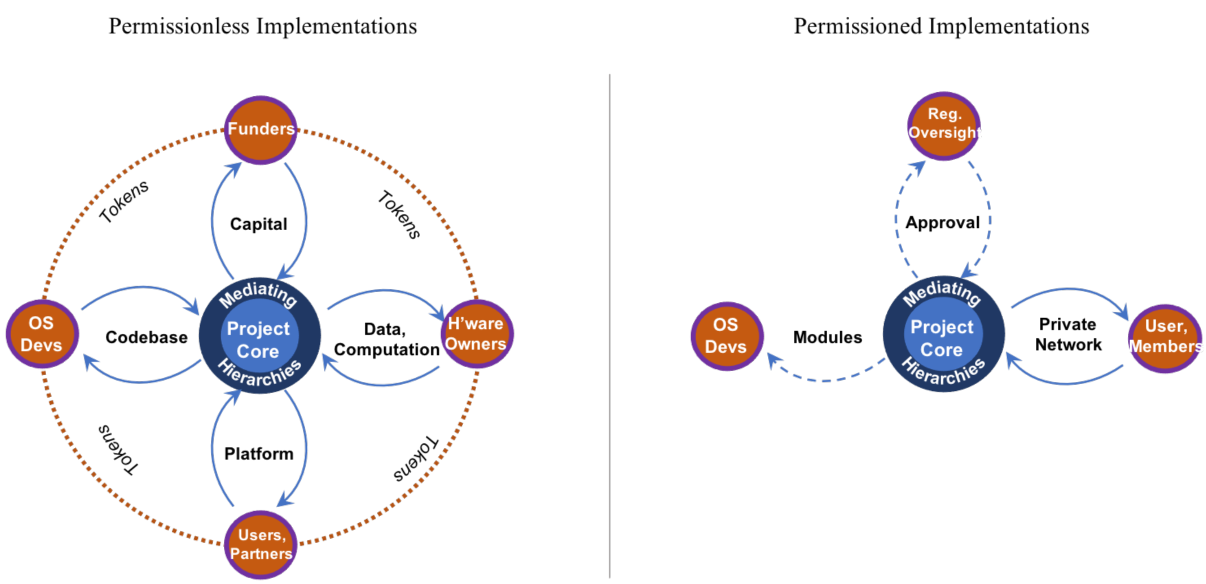
The analysis above implies the existence of two broad configurations, though these are more indicative than formal organizational types. The two configurations appear in Figure 6.5 below. The permissioned configuration to the left reflects the imperatives toward private networks and data protection inherent to these implementations, as well as the implicit reliance on regulatory approval for their viability. The resource exchanges represented by the blue arrows are in these implementations often dotted to represent some combination of either non-resource exchange (as with regulators) or attenuated resource exchange (as when minimal code is posted but not developed publicly). Boundaries in this configuration are also relatively clear, with both the open source developer community and regulators operating well outside the hierarchies and private network members, while the latter two groups are integrally linked by ongoing interaction.

The permissionless configuration to the right is more complex, with multiple internal and external boundaries. As in the permissioned configuration, the mediating hierarchies maintain a strong internal boundary surrounding the core. The permissionless configuration differs strongly, however, in its joint and ongoing mobilization of multiple external communities, and thus in its need for orchestration across and between their members' interests. The diagram shows how this work happens in two ways, one of which is direct co-creation of resources mediated by the project's mediating hierarchies (the blue arrows).

[9] https://ecf.network/

The second is more indirect but no less important and involves the use of project-specific tokens. The threat of securities regulation that followed the initial boom in Initial Coin Offerings led to widespread (and ongoing) experimentation with new types of tokens with increasingly specific uses within projects. It has since become common for projects to offer multiple tokens with differing combinations of ownership, utility, governance and other rights that are divided and allocated through token ownership. These tokens are described in project documents as central to the functioning of their overall offering, and in particular for the participation of users in their user-facing platforms. In keeping with their structure as platforms, the incentives designed into both the tokens and their use to reward contributions ideally attracted both buyers and sellers, which is to say end users and ecosystem partners.

Figure 6.5: Generalized Configurations



These archetypes in Figure 6.5 are answers to the unit of analysis problem, and in turn provide insight into the core questions of integration. The first of these is processual, and asks what holds these projects together at a given time? As noted by Puranam et al. (2014), the existence of such configurations is often evidence of the existence of complementarities. In the case of blockchain projects, such complementarities manifest in the patterned co-occurrence of resources and their modes of coordination at several key points. The configurations in Figure 6.5

point to both the presence of several such complementarities, as well as the ways in which these complementarities are the site of organizational efforts. Put another way, these complementarities are themselves key resources for the projects to manage.

The most obvious of these operate across the horizontal or technological axis, and involves the structural complementarity between code and the hardware on which it runs to create network nodes. This functional complementarity operates in theory only if it is actively orchestrated, in part through the provision of incentives and rewards to the communities providing each, as well as through the management of their intersection. This often happens through mediating hierarchies funding internal and external developer teams working on the software clients that encode node operations. Hardware owners committed to running nodes are also increasingly being brought into project-level governance decisions because of their level of commitment to the network's functioning in proof-of-stake and other formats.

The technological axis (of code and computational hardware) and is also implicated in a distinct form of layered complementarity involving the platforms of permissionless projects. As with any platform, the ecosystem of users and partners whose interaction comprises the platform are obvious structural complements, which begs the question of how that complementarity is managed. One of the key ways this is accomplished is through smart contracts, a subset of code that allows for the encoding of rewards and sanctions attached to particular actions by those in particular roles (e.g. buyer, seller, reviewer). These incentives are given force by the tokens that are attached to them, which can also assume multiple forms within a project depending on their design.

The second question of integration extends to persistence: what enables these loosely coupled structures to maintain coherence over time? The answer to this question lies in the ability of projects to evolve, a process that often involves collectively negotiating changes in core resources and project components. This evolution is central to the technology given its relative nascence, which in practice means that the protocols of active projects are not only maintained but are constantly changing and being updated as new modes of cryptography, hardware and coordination practices emerge. This constant technological evolution is a site of intensive management by projects operating alone, as well as in larger ecosystems, as the example of Golem shows. This latter pattern reflects the importance of blockchain-layer projects as shared infrastructure, and the incentives that shared aspect creates for this kind of joint production.

The final form of change reflects the fact that the decentralization so central to the ecology's ethos is recognized in practice as a process rather than an immediate goal. This commitment to what practitioners call "progressive decentralization" over time is most evident in changes to the organization(s) acting as mediating hierarchies for the project. This change is often manifested through the adoption of a DAO, or Distributed Autonomous Organization. A DAO is an assemblage of smart contracts that encode the decision-making rules, roles and functions to which its community has agreed. DAOs have grown in popularity as a means of integrating and attempting to automate coordination mechanisms, while also allowing for token-based (rather than hierarchical) participation in the decision-making process. It is important to note, however, that although DAOs are often discussed as though they embody the entirety of a project (see e.g. Hsieh et al. 2018), that is almost never strictly the case. DAOs instead are sets of encoded mechanisms that automate selected coordination mechanisms for their members. DAOs are typically introduced as projects evolve away from their early reliance on core members and become progressively more decentralized.

These latter forms of fluidity are directly implicated in changes to a project's protocol, which is to say, to the governance of each project. This combination of fluidity and control is emblematic of implementations of the technology. I address it more fully in the following chapter.

Chapter 7

Open Source Institution-Building: Blockchain Protocols and Governance

Organizational sociology has long taken coordination and control as the two primary challenges of maintaining social order. This chapter focuses on the latter of these, and asks how governance works in a distributed, digital technology.

Doing so requires both narrowing and broadening of scope. It is a narrowing in that I focus on one aspect of the technology, namely protocols. Protocols are central to the functional claims of blockchain technology to embody aspects of organizations, private legal systems, polities and economies. Once documented, typically in a white paper such as Bitcoin's (Nakamoto 2008), these protocols then circulate among stakeholders within and outside of the project, enabling coordination. Even closed or permissioned blockchain projects either still publish a protocol, or (more often) are built using existing open source protocols that are then customized. As a result, blockchain protocols are thus among the most important artifacts linking the technology and its social organization.

Understanding the role of protocols in the control of blockchain also necessitates a broadening of focus from the technology itself to the way it enrolls a wide range of actors – in other words, to governance as a sociotechnical solution to the Hobbesian question of order.

Protocols play a crucial role in the establishment and ongoing governance of distributed systems, where they are ”the technology of organization and control operating in distributed networks” (Galloway 2004, p. 317). As such, protocols-in-the-making are a site where incentives, penalties and the like are negotiated and contested, bringing stakeholders and their potentially competing interests into the discussion (Rahwan 2018).

Approaching governance from the perspective of protocols allows me to make both empirical and theoretical contributions. Empirically, this chapter joins a growing body of work that searches for mechanisms of design, coordination and governance in a range of contexts rather than focusing closely on a single case study or application of the technology (Beck, Müller-Bloch, and King 2018; Catalini and Gans 2016; De Filippi and Wright 2018; Halaburda and Sarvary 2016; Werbach 2018). The theoretical contribution is to understanding of the process of institution-building in a digitalized society.

The remainder of this chapter is organized around first developing an analytical framework, and then applying that framework. The next section provides a non-specialist overview of blockchain protocols and their construction, and is followed by a section that looks more broadly at tensions in technological governance. After describing the two-part methodological approach I developed for this study, the chapter goes on to articulate the analytical framework I use. The remaining sections apply that framework, and the chapter concludes with a discussion of findings.

7.1 Empirical Terrain: Blockchain Protocols as Recombinant Meta-Algorithms

The term *protocol* is typically used to specify a commonly accepted technological standard, such as TCP-IP for the Internet (Galloway 2004). I define the term *blockchain protocol* more specifically to mean *the core technological logic, incentive system and code that jointly define the functionality of a blockchain project.* In contrast to the traditional definition of protocol, blockchain protocols are not yet institutionalized as widespread standards but rather are meant to define the functioning of individual networks and ecosystems, as well as providing a point of technological intersection with other networks and ecosystems. Their proliferation and competition are to be expected in this early stage.

The central role of code in blockchain protocols gives them a clear affinity to algorithms, though that term implies a unity of structure and purpose that doesn't fully capture the sociotechnical complexity of blockchain protocols. This complexity in part reflects the incorporation of both encoded and social rules into the protocols. It also results from the fact that these protocols are constructed from a set of modules. As a result, blockchain protocols are more aptly described as *meta-algorithms*. This characteristic embodies the paradox at the heart of open science and technology in that the components of protocols are in simultaneously technological instruments and 'knowledge objects" (Knorr Cetina 1997) under constant revision, even as the protocols themselves are "never quite themselves" (ibid, p. 13) but rather both subjects and objects of expert development.

The four generalized components used by designers of these meta-algorithms are as follows:

- **Base protocols.** Base protocols are typically published in academic or quasi-academic open access repositories in computer science and related disciplines. Most consensus algorithms begin as base protocols (e.g. variants of Practical Byzantine Fault Tolerance [Castro and Liskov 1999]). While primarily algorithmic, these base protocols often also include (in the parlance of the field) "social" rules that do not appear in their code. The most famous example is Bitcoin's use of the longest-chain rule, which features prominently in its white paper but is practiced as a social incentive, rather than an algorithmic one (Nakamoto 2008).

- **Cryptographic primitives.** These building blocks include algorithms for hashing and digital signatures and can also include other security measures such as zero-knowledge and other forms of proofs (Wang et al. 2019). This work often begins as academic research before diffusing to practice, though the lines between the two are often blurred. In other cases, such as cryptographic hashing functions, the function may have been developed as part of a global contest sponsored by an agency such as the National Institute of Standards and Technology, with the winners chosen for the level of security provided by their submission.[1]

- **Cryptoeconomic primitives.** These primitives include cryptoeconomic mechanisms linking tokens to incentive design. Examples include coded mechanisms for staking, delegation and network fees.

- **Existing code libraries.** Nearly every component in this list exists in a code library on Github, whether as a standalone implementation or as part of the reference implementation of an existing protocol. Existing code libraries are a primary input into the development of blockchain protocols.

As a generalization, blockchain protocols are typically constructed out of components selected from the available options in each of these four categories of components. The decision about which components to select also necessitates decisions about specific parameters to use within each. For example, the decision to use one form of consensus algorithm necessitates a decision about the safety tolerance threshold for misbehaving nodes (one third? one half?). The design and ongoing evolution of blockchain protocols is thus - at its core - a series of decisions about components and parameters within those components.

Given the nature of the modules on which they draw, blockchain protocols are also inherently part of open science and open source code development. This is in part because of the relative youth of the technology. These modules are part of active research and development of consensus and other base protocols by academics, practitioners, government agencies and corporations. What keeps this from being a chaotic free for all? Protocol designers are expected to use cryptographic and other functions with known risk parameters and credibility established by surviving ongoing attempts at breaking their security. This is known colloquially by the phrase “don’t roll your own crypto” (attributed to Schneier 1998). This stance reflects the scientific conventions of skepticism and verification (Owen-Smith 2001), along with the risk mindset of

[1] NIST undertakes a broad spectrum of projects in cryptography, security and related disciplines. See https://csrc.nist.gov/projects.

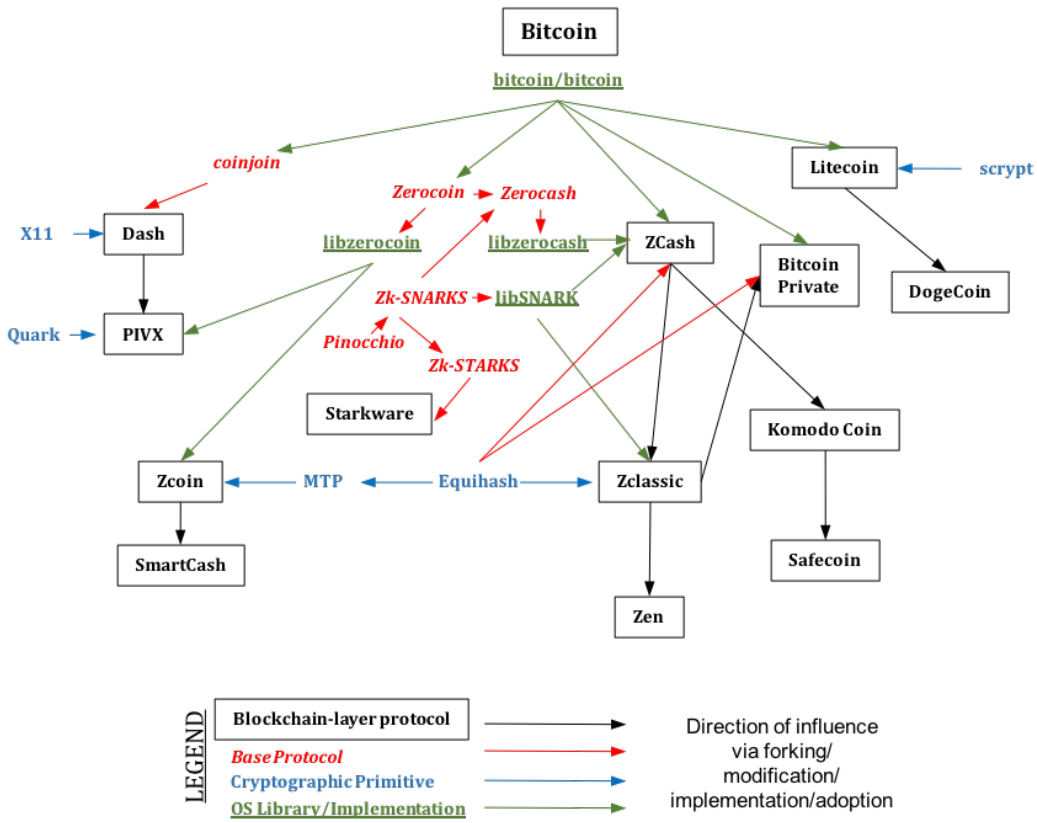
cryptography as a form of security. As a result, it is rare for a blockchain protocol to premiere an entirely new component in its construction.

Blockchain developer communities are also deeply enmeshed in the practices and structures of open source code development (Brunton 2019; Swartz 2018). Open source communities operate under rules specifying more or less complex rules or architectures of participation that define contributing, debating and ultimately voting on contributions (Baldwin and Clark 2006; Geiger 2017; West and O'Mahony 2008). At their most general, these architectures define a process of *propose, deliberate and decide* in which the processes of each stage and the progress between them are defined according to the standards of the community while broadly following various implicit templates. Technically, anyone can propose a change to the code, but in practice these processes establish and enforce standards that set a high threshold of technological and project-specific knowledge for meaningful participation.

Figure 7.1 gives a stylized view of historical trajectories of recombination based on Bitcoin's core code repository or "repo", and shows how it has been recombined with other innovations over time in the design of later protocols. The core Bitcoin repo, called "bitcoin/bitcoin" appears in green, as do the other code repos for other projects that form part of the trajectory. Items in red are base protocols, and enter the chart at points where they are merged into new protocols. Items in blue are cryptographic primitives (primarily hashing functions) and are shown similarly at their point of incorporation into the blockchain-layer protocols that appear in black text boxes.

The trajectory shown in Figure 7.1 represents innovations developed in response to the perceived privacy shortcomings inherent in Bitcoin's blockchain. The earliest efforts in this vein, such as coinjoin (Maxwell 2013) and Zerocoin (Miers et al. 2013), were designed as additional layers for Bitcoin itself based on the functionality of Bitcoin's core code library on Github, namely bitcoin/bitcoin. Dash, an early and influential codebase fork of Bitcoin, adopted and modified coinjoin in order to deliver enhanced privacy, in part by incorporating the X11 hashing algorithm. While coinjoin was created by a practitioner, Zerocoin was created by a team of academics who later revised it to include innovations in zero-knowledge proofs, and named the newly revised

Figure 7.1: Assembling Blockchain-Layer Protocols



Sources: Github, project white papers.

version Zerocash. The former protocol remains influential for several cryptocurrency projects, while the latter became the basis for Zcash, among the most influential cryptocurrencies given its technological advances. The team of academics who developed Zerocoin and Zerocash have since created a new version of zero-knowledge proofs called "zk-STARKS", which is the basis of their new commercial project named Starkware Industries, which has its own Starkware protocol.

7.2 Tensions in Technological Governance

Governance institutions have been a central concern for the designers of blockchain technology from its inception, with the very first or "genesis" block of Bitcoin's blockchain inscribing a critique of the traditional financial system it sought to circumvent: "The Times

03/Jan/2009 Chancellor on brink of second bailout for banks." This combination of critique of existing governance institutions and the pursuit of technologically and politically superior alternatives reflects the deep roots of both Bitcoin and the larger field of technology it inspired in the emancipatory commitments of the cypherpunk movement from which they grew (Brunton 2019; Swartz 2017). Timothy May's famous manifesto for the movement referenced several technological advances he foresaw as central to its objectives that anticipated components of blockchain, including "methods [] based upon public-key encryption, zero-knowledge interactive proof systems, and various software protocols for interaction, authentication, and verification" (May 1992).

The ongoing development of these and other technologies in distributed systems, cryptography and other disciplines created the building blocks for Bitcoin's design (Narayanan et al. 2016), which was one effort among many to create digital alternatives to fundamental institutions including money, contracts, property rights, law, voting and accounting (Back 1997; Dai 1998; Grigg 2004, 2005; Merkle 2016; Szabo 1997). These institutions are notable for their role in enabling the shift from interpersonal to large-scale exchange during the industrial transformation (Zucker 1986). The designers of the alternatives to these institutions sought to develop technologies that maintained or improved on the functional characteristics of the original institutions. They sought to do this by replacing the authority of the centralized state with distributed alternatives based on a combination of cryptography and distributed computation. As a first approximation, cryptography was introduced to replace the authority of the state with that of mathematical proof, while distributed computation replaced the centralized functioning of the state.

Concerned from the start with institutions, the designers of blockchain will realize their visions only to the extent that they can create effective digital variations on core institutions of governance. Two closely related aspects of distributed sociotechnical organization are relevant to that goal. The first is the potential use of technology for decentralization. Decentralized or distributed technology capitalizes on the ability to build networks of linked devices acting in concert. The second involves the democratization of social processes by opening them to broad and meaningful participation. Decentralization and democratization have obvious theoretical affinity as political concepts. Literature on governance across the social sciences has for decades theorized a transition away from Hobbesian leviathans of the state to more decentralized structures such as markets and networks (Klijn and Skelcher 2007; Levi-Faur 2012; Powell 1990; Sah and Stiglitz 1986).

Theorists and developers of decentralized technology have emphasized its democratic potential to lower or remove barriers to participation, whether by providing an alternative to traditional leviathans or by creating entirely new modes of organization. Democratization has been studied extensively in the context of open innovation and open source code development, where the “crowd” or community is the source of ideas, knowledge, and problem solutions (Benkler 2006; von Hippel and von Krogh 2003). Discursively, democratization is typically evident as an assumed set of affordances of “anyone can...,” whether that means running a network node, contributing to opens source code or interacting without intermediaries. These claims echo those made in the early days of the Internet, when freedom was linked to the liberating affordances it offered as an open, generative platform for participation in the information economy (Benkler 2006; Castells 2011; Zittrain 2008).

The Internet’s subsequent evolution, however, provides a cautionary tale that illustrates the contingency of any relationship between decentralized technology and democratization. The Internet we have today - with access for the vast majority of users governed (whether formally or informally) by a shrinking set of increasingly unaccountable monopolies - functions for the vast majority of its users as a very different structure than was envisioned by its early champions (Benkler 2016; Lanier 2018; Zuboff 2019). Zuboff’s work shows that this outcome was neither inevitable nor necessary, and was instead the contingent result of a series of governance decisions that had the effect of redistributing knowledge, authority and power (Zuboff 2019).

The arrival of blockchain as another set of technologies that valorize decentralization as a simultaneously technological and political project invites a sociotechnical analysis of governance that engages with issues of power, politics and technology as constitutive choices rather than fundamental properties. This offers an opportunity to revisit core notions of governance in a new digital context in which the subject is a central concern. In particular, it casts the central question of decentralized governance in a new light: how is a decentralized technology developed to assume some functions of centralized institutions governed? How does technology that adopts institutionalized functions govern, and how is it governed?

Blockchain protocols are a particularly apt research site for answering these questions. As artifacts, protocols offer insights into the tensions between participation and control. To the extent that tensions in design and governance persist - centralized vs. decentralized, controlled vs. democratic - they will be worked out in the context of these protocols and the governance decisions with which they are imbricated (Mumford 1964).

Two tensions are particularly germane. The first of these concerns the tension between individual participation and scale. Variants on the relationship between local action and scale appear in theories of both democratic decision-making (Arrow 1950; Fishkin 2011; List 2011) and distributed systems (Gilbert and Lynch 2002, 2012), which formalize tradeoffs between desirable design goals (e.g. participation) and scale. This work is often typified by the development of “impossibility theorems” that establish the necessity of such tradeoffs given the impossibility of achieving all of a set of design goals simultaneously.

Zuboff’s analysis of the political economy of incumbent platforms such as Google and Facebook points to algorithms as a technological solution to this challenge of scale. In this framework, algorithms are both products of governance decisions as well as shapers of social orders within the bounds of the systems they enable. Approaching blockchain protocols through the lens of algorithms offers an opportunity to look beyond code to the coders and other actors who design and exert influence over protocols as important governance actors (Grimmelmann 2005; Fourcade 2017; Seaver 2017).

The second tension embodies one of the oldest debates in Western political theory (Caramani 2017), namely the tension between governance by experts and more participatory governance.

This tension is particularly acute in settings where technological complexity limits meaningful participation by the broader public, an aspect that has also led to critique of its anti-democratic tendencies ((Borrás 2012; Michels 1966) Even small, intentionally "structureless" groups are prone to domination by a few members when they include subgroups with social ties, similar ways of thinking and dense informal communication (Freeman 1972).

Open-source development practices reflect this pattern by tempering democratic aspirations with incentives for participation by developers with demonstrated skill. Similarly, blockchain protocol development and governance are often developer-centric, though there is a strong tension between this tendency and the technology's explicit use of economic incentives that enroll several other types of actors. This tension is among the most generative sources of both governance mishaps and of innovation in the overall ecology.

7.3 Data and Methods

The overarching challenge of studying a phenomenon in medias res would seem to be one of identifying data without the validation of history. In reality, the combination of open technology and experimentation at the heart of blockchain technology means that the primary challenge is instead one of "excess" (Abbott 2016). In addition to the proliferation of thousands of implementations, there are up to a dozen or more venues where project members describe their projects' ongoing evolution, state and revise objectives, and engage in debate. Arriving at generalizable conclusions from this wealth of data necessitated some means of identifying key variables and then tracing those variables across the ecology of projects.

I took a two-stage approach to the challenge of shaping this excess into a tractable analysis. The first stage was to build an analytical framework to scaffold and direct the analysis by identifying important variables and levels. I ground this integrative framework in generalized variables and insights from Elinor Ostrom's Institutional Analysis and Development (IAD) framework (Ostrom 1990, 2003, 2005), which I then customize for my analysis using key findings regarding aspects of governance related to algorithms, institutions and open source. Ostrom and her colleagues are perhaps best known for their extensive case studies of the localized governance of common-pool resources such as waterways and fisheries in the absence of state hierarchy and private property (Ostrom 1990). This chapter, however, builds on Ostrom's later work, in which she developed a more generalized framework for the analysis of institutions as a solution to collective action problems including – but not limited to – those involving common-pool resources (Olson 1971; Ostrom 2003, 2005, 2017). The Institutional Analysis and Development (IAD) framework Ostrom developed over the course of her career (and articulated most fully in Ostrom [2005]) explicitly recognized the diversity of institutions. Rather than seeking a single model or theory, it sought instead to identify what she described as the universal building blocks of institutions and to define them as variables within a general framework. I take these building blocks as a starting point for describing institutions, and use them to adjust and revise the analytical framework for my own analysis by incorporating aspects of algorithmic and open source governance.

The analytical framework provided the scaffolding for the second stage of my analysis, which involved populating the framework's variables and levels through an analysis of the protocols of major projects. I undertook this second stage of analysis using techniques of algorithmic ethnography (Christin 2020; Seaver 2017). This sociomaterial approach rejects the notion of algorithms as unknowable black boxes and instead treats them as parts of larger assemblages of diverse actors and code that are under constant construction (Seaver 2017, 2019). This constant change enrolls different actors in different facets and stages of the algorithm's development and use, and necessitate a broader analytical lens that incorporates this change and works across available sites and data (Christin 2020). I draw on strategies of scavenging, or gathering material from as many sites and data types as possible, as well as comparison and triangulation across these sites and data types (Christin 2020; Seaver 2017). My primary data sources here were a combination of Github, project-level blogs and white papers, and online fora such as Reddit and project-level discussion forums.

7.4 Analytical Framework

The generalized IAD framework explicitly builds on theoretical contributions from across the social sciences. As a methodological individualist, Ostrom drew from game theory a concern with defining the rules-in-use governing individuals' choices; she took from economics and law an understanding of property rights as a "bundle" to be distributed differentially to players based on the rules of the game. She defined these games in sociological terms of situated action as "action situations" or arenas shaped by a combination of the broader community's shared expectations as well as a set of governing rules-in-use. Those rules specify both the actions available to actors as well as the outcomes linked to those actions and are nested in levels that reflect the nature of the governance decisions linked to the establishment and evolution of rules in use. By incorporating multiple levels, Ostrom argues, institutional analysts can identify the mechanisms and structures that operate both within and across levels and can more effectively account for the relationship between local decisions and large-scale outcomes (Ostrom 2005).

Analytical Variables

In this framework, a functioning governance system clearly defines both ***roles*** and the specific ***rights*** allocated to each (Ostrom 2003; Ostrom and Schlager 1996; Schlager and Ostrom 1992). In Ostrom's early work on natural resource governance, these rights varied among roles deriving from different aspects of access, use and decision-making. Ostrom's later work on man-made information commons in open science introduced additional categories for contributors (Hess and Ostrom 2003, 2007). Delineating these roles is particularly germane in the domain of platform-based ecosystems, where efforts to build theory have focused on the tensions between fostering open innovation and seeking some measure of control over its course (Adner 2017; Wareham, Fox, and Cano Giner 2014).

The rules in Ostrom's framework specify not only positions and associated rights, but also the rewards or sanctions that shape behavior. The nature of these ***incentives*** comprise the second analytical variable in the framework. In the 21st century, the nexus of incentives online is largely driven by research at the intersection of economics and computer science that blends advances in the economic design of markets (Roth 2002, 2008) with computational advances in algorithmic game theory (Jackson 2013, 2014; Nisan et al. 2007; Roughgarden 2010). The basis of this work is the use of game theory to model and create incentive mechanisms for individual behavior that aggregates to beneficial outcomes. Rather than imposing sanctions, much of this work relies on creating incentives for individuals to converge on the same decision. This mode of encouraging convergence is grounded in extensive research into the tendency of individuals to coordinate around shared points of reference or so-called Schelling or focal points (Schelling 1981; Sugden 1995).

The importance of incentives is often sharpened by their absence. Although the organizational literature increasingly addresses questions of governance and emergent structures in open source, it has largely ignored the meso-level issues of incentives necessary for maintenance of these systems.[2] Enterprise software development in the 21st century relies heavily on infrastructural, open source libraries or modules of code that are maintained by devoted - but largely uncompensated - developers (Eghbal 2016, 2020). The growth of this form of infrastructure has made issues of incentives, and in particular the lack of structural and reliable funding for the maintenance of heavily used open source code libraries, a central concern for the ongoing viability of open source.

The third analytical variable captures the centrality of rules for cooperation and collective decision-making in Ostrom's work on governance. I follow Ostrom in calling these ***rules-in-use***. Together with roles and rights, rules-in-use are perhaps the primary distinction among the three levels I use for my framework. The rules-in-use for my analysis include the decision rules for collective choice for which Ostrom is best known. In addition, at the operational level, operational rules guide behavior in everyday situations, while separate rules are triggered in the event of a lack of social consensus on the best path forward (Elster 1998).

[2] Classic early studies of open source software addressed these incentives but situated them at the micro level of individual motivation (Lerner and Tirole 2002, 2005) and preceded its subsequent transition from being an exception to traditional economic activity to become foundational to it.

Ostrom's work has from the beginning sought to answer the question of how groups can develop their own, localized institutions as solutions to collective action problems. Those institutions are in turn rooted in the collective expectations of their communities, which provide a necessary basis of legitimacy for the cooperation that the institutions enable (and for the development of the institutions themselves). These expectations are the result of ongoing work by community members, whose expectations form a social consensus on values and norms that defines appropriate behavior and establishes a basis for legitimate authority (Blau 1963). I term this final variable ***community consensus.***

Tensions between open participation and specialized knowledge are particularly salient for this variable given the need to define the boundaries of the community making decisions. Administration by experts has been a primary mode of governance for science and technology, where the consensus of experts provides an autonomous source of legitimacy (Merton 1942; Starr 1982). This approach is evident in the evolution of "rough consensus" as a mode of deliberation in open source software development. Rough consensus as a concept was originally formalized by the Internet Engineering Task Force (IETF), a global organization of scientists and researchers working on various aspects of the Internet protocol and its related standards. The IETF's definition of rough consensus combines a commitment to enfranchisement and open deliberation with assumptions of (and implicit deference to) the technical expertise of those participating.[3] It is participatory in the sense that anyone present can voice their opinions, but antidemocratic in its avoidance of pure majority rule and the attendant risk of silencing useful objections.

Political theorists have approached the issue of consensus-building as more of a bottom-up and participatory question of deliberation, which draws its legitimacy from justification and accountability. Theorists of deliberation frame it not as a replacement for decision-making but rather a means for attempting to build a social consensus that justifies policy decisions and creates accountability for decision-makers (Habermas 1989; Rawls 2005). Some historically minded analyses focus on broader structural options at moments of "constitutive choice" (Starr 2019)

[3] See Request for Comment (RFC) 2418, https://tools.ietf.org/html/rfc2418, later refined in RFC 7282, https://tools.ietf.org/html/rfc7282.

when normal politics are no longer adequate, and the underlying framework needs to be modified. At such moments, popular sovereignty lies not in the vote, but in "arguing, mobilizing, recruiting a broadening commitment" (Ackerman 1998, p. 4).

Structural Dimensions: Polycentric Resources and Governance

Vertical nesting of decision levels is critical to Ostrom's conception of multi-centered governance. She distinguishes three levels of decisions—the operational, collective choice, and constitutional—that increase in breadth of impact as the analysis moves deeper into the governance structure. This multilevel structure reflects increasing entrenchment (Starr 2019) involved in decisions that involve growing sets of stakeholders. In what follows, I relate Ostrom's three levels of decision situations to the broader literature on code, algorithms and governance.

In the first of the three levels, *operational* situations, rules-in-use shape the day-to-day behavior of participants based on their roles, incentives/rewards and other domain-specific factors. When operational rules-in-use are encoded into algorithms, they result in algorithmic governance, a term that has come to imply a form of control exerted through algorithms on behalf of others (Kellogg, Valentine, and Christin 2019; Lessig 2006; O'Neil 2016; Pasquale 2015, 2019). A rapidly growing literature documents the increasing prevalence of algorithmic techniques in both the private and public sectors as a new form of social control that shapes individual action (Citron 2007; O'Neil 2016; Pasquale 2015). I take this approach in Section 5 of this chapter in examining *operational governance by protocols* in blockchain projects.

Most practitioners mean something different than algorithmic governance when speaking about "blockchain governance" (Buterin 2017; Ehrsam 2017; Prewitt and McKie 2018; Zamfir 2018). In these settings, code and protocols are the object rather than the means of governance, and the focus is on decision processes for making changes and upgrades to protocols; in the language of governance, this is the governance *of* protocols, as distinct from governance *by* protocols (DeNardis and Hackl 2015). These decisions unfold at a deeper layer, which Ostrom would define as *collective choice* situations, that is, settings in which communities change their operational rules.

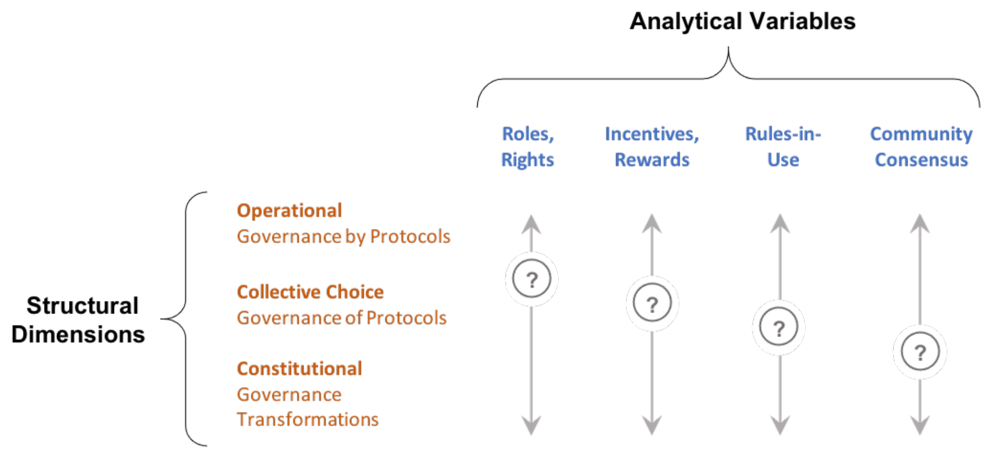
The deepest, most entrenched level of decision-making is what Ostrom terms the *constitutional* level, which determines “who decides who decides” (Zuboff 2019, p. 181). Decisions taken here are among the most consequential in that they result in changes to the structure of governance that ripple through the other two levels. These can include changes to collective choice rules, the establishment (or amendment) of a constitution, and the creation of a new protocol. In traditional institutions, the powerful - and potentially destabilizing - effects of these decisions have engendered mechanisms that prevent them from happening too frequently, making them a kind of exceptional politics undertaken only when the “normal politics” of existing collective choice rules fail (Starr 2019). In order to distinguish this level from changes to protocols such decisions can entail, I define such *constitutional transformations* as decisions that involve significant changes to any of the four analytical variables I describe above.

Framework Summary

The framework I employ for the empirical analysis that follows flows directly from the combination of analytical variables and vertical, structural dimensions I establish above. While it draws inspiration from Ostrom’s analysis, it also incorporates insights from more recent work on algorithmic governance, particularly in my articulation of the three levels of governance. In keeping with Ostrom, I frame these three levels as nested, which is to say, as related hierarchically in their effects. From the bottom up, the levels establish a hierarchy of control in which each level defines the nature of the level(s) above. From the top down, the levels establish a hierarchy of change, with tensions at a given level only resolved at deeper levels.

This hierarchical, nested framework provides the structure within which I explore each of the four analytical variables. While each appears in each of the three hierarchical levels, it is not obvious ex ante how each functions at any one level. I use these variables to shape my analysis of the mechanisms of governance specific to each level; this also allows me to refine my later analysis of the hierarchical interactions between levels. Doing so requires filling in the conceptual framework in Figure 7.2, below.

Figure 7.2: Analytical Framework



7.5 Level I: Operational Governance by Blockchain Protocols

By establishing the positions, rules and rewards/sanctions that shape behavior in blockchain networks, protocols act analogously to Ostrom's operational rules-in-use. I explore each of these aspects below.

7.5.1 Roles and Rights

One of the most striking aspects of blockchain protocols is the ontological uncertainty surrounding the status of the *user*, whose participation is typically left unspecified or conflated with running nodes in the network. This vagueness is clearest in the original Bitcoin white paper, in which users are mentioned twice while nodes are mentioned 39 times using agentic language (e.g. "nodes can leave and rejoin the network at will," or "each node collects new transactions into a block) (Nakamoto 2008) . This simplicity reflected the decade-old conditions of technology and participation at the time of Bitcoin's founding. The original vision of Bitcoin was that every user of the system would also be a participant in the system, and that these users would all run the same software and contribute a part of their computers' capabilities to its maintenance. The success of Bitcoin beyond its original vision as a payment system led to the first splitting of these roles as investors began to buy Bitcoin. Demand from investors raised the token's price, increasing the rewards for mining, and thus created incentives for those interested in mining to invest in increasingly specialized and expensive equipment, expanding the categories of stakeholders beyond those envisioned in the initial protocol, and laying the groundwork for failures of social consensus I describe in subsequent sections.

Both the roles of users as well as the types of nodes involved in blockchain systems have grown more complex since Bitcoin, largely because of a sea-change introduced with the launch of Ethereum's initial or beta network in 2015. Where Bitcoin was designed purely as a means of facilitating peer-to-peer payments, Ethereum was designed with the intent of providing a generative platform for the development of third-party applications and systems ranging from crop insurance to separately tokenized economies. In addition to providing security and decentralizing control, the peer-to-peer network in Ethereum was envisioned as providing a global virtual computer whose aggregated processing power would support a wide range of activities grounded in computation.

The Ethereum white paper also built on and refined a distinction in the nature of nodes in the network that first appeared in the Bitcoin white paper. Although the latter's discussion of nodes broadly focused on those nodes that would participate in the network's full functioning of transaction validation and block consensus (on which more below), the Bitcoin paper also identified a set of nodes with less responsibility to the network and thus less functionality. These ”Simplified Payment Validation” or SPV nodes would interact with the information-rich nodes in the core of the network but only to process transactions. Ethereum made a similar distinction between full nodes, which carry out the core functions of the network and require the greatest commitments to providing processing power and connectivity, and light nodes, which can read from the network but typically only connect to full clients in order to get their information. This distinction between two or more node types and functions has become the norm for blockchain-layer protocols. It reflects the often unintended introduction of hierarchy as a means of taming complex decision processes.

Another development has been the rise in importance of token "stakeholders," defined narrowly as token owners willing to bond or "stake" their idle tokens in order to participate in the network. Proof-of-Stake consensus protocols limit participation in the pool of consensus nodes to those whose holdings exceed a protocol-defined threshold, with a stakeholder's likelihood of being chosen rising with the size of their stake. This approach has the effect of enshrining a property-based version of franchise, with its attendant tendency toward capital-based concentration. Variants on Proof-of-Stake add a role for stakeholders opting to delegate their participation to other stakeholders. In some cases, such as Tezos and Cosmos, delegating one's tokens to a given delegate can also earn a token holder a portion of the fees earned by that delegate, giving token holders more of a stake to participate in voting and monitoring.

The Polkadot protocol is indicative of the expansion of roles in protocols since Bitcoin. Polkadot is a combination of a blockchain-layer protocol (based on its sub-protocol, Substrate) as well as a means of integrating and working across multiple independent chains. As a result, the primary function is validating transactions on the independent chains (called "parachains") and the main or relay chain, though it is only one of four named roles in the protocol.[4] Stakers are token holders who want to participate in the consensus process but lack sufficient holdings to become Validators, who undertake the responsibility of validating blocks of transactions from parachains as well as participating in the consensus process for the main or relay chain. Nominators have the right to participate indirectly by nominating Validators. Collators translate between parachains and Validators by checking and bundling transactions into blocks, and Fishermen monitor the entire system.

7.5.2 Operational Rules-in-Use

The core of operational decision rules in blockchain-layer protocols lies in the methods by which they bring the autonomous nodes in the network into agreement. These methods are

[4] https://github.com/paritytech/polkadot/wiki/Polkadot-Roles-&-Actors

typically gathered into a single consensus mechanism or algorithm. Consensus in distributed systems is technically defined as the outcome of a process by which the nodes in the network reach agreement on the state of the world in an uncertain environment van Steen and Tanenbaum 2017). The consensus algorithm is central to the ability of blockchain-layer protocols to deliver the security and finality of transactions without the intervention of one of the trusted institutions or intermediaries studied by Zucker (1986). Without this ability to create a single authoritative history, neither the tokens of individual blockchains nor the blockchain-enabled protocols that build on them would be viable.

One of the first attempts at a consensus protocol took as its organizing metaphor communications between a distributed network of individual parliamentarians (as a metaphor for nodes in a networked process) and the challenge at arriving at a global agreement on the state of legislation regarding war (as a metaphor for the overall state of the distributed, or decentralized, system and its ledger) (Lamport 1998). Lamport's colorful metaphors of "Byzantine" or arbitrarily untruthful generals was adopted as the Byzantine Generals Problem and formalized into a general definition of a threat model in terms of the proportion of the network nodes that must be trustworthy in order for the blockchain to function properly. For example, the Practical Byzantine Fault Tolerant consensus algorithm allows for up to one third of the nodes in the network to be malicious actors (Castro and Liskov 1999), while Bitcoin's protocol allows for up to one half.

In keeping with their algorithmic nature, blockchain-layer consensus mechanisms are designed to coordinate decision-making in a way that attenuates the proposal-deliberation-decision process in favor of automated rules governing the proposal and selection of a new block of transactions as the next definitive one in the chain, as well as the selection of the authoritative version of history to which to append them. The most widely studied consensus algorithm is Bitcoin's approach, in which full nodes compete for the right to produce the next block by providing evidence of having performed expensive computation or "proof of work." The first full node to do this is the winner. Once this authoritative block is identified, the next question is what chain to attach it to. The rule in Bitcoin (widely but not universally followed in other blockchain-layer protocols) is that the chain that is the longest, and thus represents the most aggregated computational work, is the one. While this remains an important algorithm in the field, especially among cryptocurrencies, it is by no means the only approach. An increasingly popular alternative is Proof-of-Stake, in which the operators of full nodes first bond or stake some amount of tokens and are then selected to produce the next block based on a weighted random function the size of their stake. The most widely used variant of Proof-of-Stake involves delegation by stakeholders to a subset of nodes. I describe this variant – Delegated Proof-of-Stake – more fully below.

7.5.3 Incentives

At the level of blockchain-layer protocols, the work of coordinating decentralized activity draws heavily on economic mechanism design and cryptographic methods that create incentives and (when possible) sanctions for the behavior of full nodes while also maintaining security. Much of this work is oriented to building clear points of coordination (Schelling 1981; Sugden 1995) for cooperating actors while also deterring misbehavior that would damage the system. The primary goal of such coordination is to ensure that nodes can quickly and reliably settle on the single authoritative version of the blockchain with each new block, and that malicious actors are either unable or unwilling to interfere in this process.

While these challenges have long been inherent to the design of distributed systems, the addition of programmable tokens specific to each blockchain expands the space of possible designs by adding monetary incentives and sanctions that are denominated in a currency whose value is (in theory) also tied to the health of the network. Bitcoin pioneered this blending of monetary and algorithmic design. Bitcoin's designer(s) Nakamoto introduced the metaphor of gold mining as a physical analogy for the process of performing computational work that secures the network in return for a reward of money created or minted to compensate for that computational work. These "mining rewards" serve a double purpose in that they both encourage those with hardware to commit it to the network in return, and also provide a controlled schedule for the creation of new Bitcoins according to a schedule established in the protocol. The success of this approach is evident in the sheer amount of "mining" power applied to securing Bitcoin, an enormous amount of computation that has grown along with the value of the cryptocurrency. The protocol's requirement that at least 50% of the nodes behave honestly translates into a rough hurdle that any attacker would in theory need to be able to muster more than half of the aggregated computation applied to securing the network, an increasingly difficult proposition as this computational power has grown.

The growth in both the token's value and the aggregated computational power it incentivizes are also important to the creation of Schelling or coordination points. These points are crucial to the process of maintaining a single, authoritative record of transactions, which they encourage by providing incentives for the entire network to converge on a single chain. In Bitcoin (and several other blockchain-layer protocols), this is established through a rule that the longest chain of blocks - meaning, the chain that represents the most computational work performed over time - is the authoritative one. The incentive here is indirect, in that any miners working on other versions of the chain will risk wasting their efforts by mining for a blockchain with little economic value.

Bitcoin's linkage of money creation to rewards for miners laid the groundwork for a subsequent expansion in both the variety of approaches to incentivized block production (as described above) as well as the potential to link a larger number of roles and actions to economic incentives. Although these vary and are an area of rapid scientific development, these algorithms share the characteristic of requiring some form of commitment to the network. As mentioned above, Proof-of-Stake makes the bonding or staking of significant amounts of tokens the baseline for participation as a node in the consensus process. Unlike mining in Proof-of-Work, staking requires putting some portion of the staked tokens at risk, creating the potential for sanctions (called "slashing") in the event the staker departs from the protocol's rules.

The steep requirements for always-on, high-volume data connections and massive but specialized computational power for those running nodes often exceed what is readily available to individuals. As a result, these newer variants reflect the fact that they were developed in a very different environment than the ”one CPU, one vote” envisioned at the outset of Bitcoin, though the possibility of developing means of reaching consensus using non-specialized machinery has motivated an ongoing arms race between protocol and hardware developers. One solution to this challenge is evident in blockchain systems using Delegated Proof of Stake or *DPoS* as their consensus mechanism. In this structure, individual token holders vote for a small subset (by number) of agents who then participate in the process of creating and validating new blocks. Given economies of scale and the complexity of the technology, together with the potential for profit from mining and transaction fees, these delegates are typically well-funded efforts that compete in ongoing elections for votes from token holders.

7.5.4 Community Consensus

Given their algorithmic nature, blockchain protocols define consensus narrowly within their normal functioning as an outcome of a well-functioning technological process. This replacement of the messy process of reaching community agreement (called “social consensus” in the field) with automated rules for machines is fundamental to the claimed ability of the technology to create authoritative records that are difficult to tamper with.

This is not to say, however, that agents operating in these systems operate free of social conventions and requirements for agreement. Mechanisms such as Bitcoin’s longest-chain rule rely on an engineered form of consensus for actors that relies on incentive design to lead participants toward a common behavior. By contrast, purely social consensus has historically been developed and manifested outside of these protocols, though it is of crucial importance to collective decisions about changes to the protocols themselves. It is also increasingly being incorporated through experimental approaches that bring voting into the functioning of the protocols themselves. I discuss these further below.

7.6 Level II: Collective Choice Governance *of* Protocols

Blockchain protocols are constantly evolving as both the underlying technologies and the sciences of distributed systems and cryptography advance. Their designers also continually experiment with and incorporate new features and expansions in functionality. As Ostrom recognized, this process of change is qualitatively different from the daily functions encompassed in operational rules because it requires reconciling the potentially divergent interests of a broader group of stakeholders. For blockchain protocols, these stakeholders extend well beyond consensus participants to include token holders, developers, application developers and others. The question of how these stakeholders get to participate, and in which stage of decision-making, is the central challenge of collective choice governance of protocols. Success often hinges on the ability to enroll these stakeholders in the process of arriving at *social* consensus on highly *technical* issues.

7.6.1 Roles and Incentives

While there is wide variation in terms of specifics, the most common roles for individuals in the governance of protocols are generally threefold. The first of these are *developers* who are involved in maintaining and revising the core code and functionality of the protocol. This group is typically subdivided into a core group of developers or "core devs" and smaller teams dedicated to particular parts of the protocol. The second group are the *provisioners* of necessary resources for the network's functioning, typically those who operate nodes (meaning, those who own or rent hardware and run the protocol's code on that hardware). The interaction between these two groups is perhaps the most fundamental determinant of the evolution of a protocol over time given the decision rules that govern their interaction. A third group - *users* - are most often gestured to in the context of their ability to build applications on a protocol (implying a distinct set of *developers*) but are inconsistently included in governance relative to *developers* and *provisioners*.

Open source foundations are the central organizational actors in open source software projects. Prior studies have emphasized the role of open source foundations as boundary organizations that keep code and commerce distinct by housing the intellectual property rights and negotiating contracts with external actors (O'Mahony and Bechky 2008).[5] In recent years, however, open source foundations have expanded their role from buffering core contributors to cultivating and orchestrating larger ecosystems of projects. The Linux Foundation, for example, describes itself as "support[ing] the creation of sustainable open source ecosystems by providing financial and intellectual resources, infrastructure, services, events, and training."[6] This function gives such foundations two additional roles, the first of which is *funder* of development efforts. The second, closely related role is that of *network orchestrator* to the extent that the foundations use both their funding and other means of coordination to shape the development of larger ecosystems of related projects without assuming traditional forms of hierarchical control (Dhanaraj and Parkhe 2006).

As with open source development more generally, non-profit foundations in blockchain are typically created to act as third-party, independent holders of intellectual property (IP) rights in protocols and their related names, and also tend to fund and coordinate much if not most ongoing development directly. Unlike open source, foundations are typically joined in this role by independently managed for-profit firms that also provide funding and other services. These for-profit firms are most closely engaged as well-funded *provisioners* of costly network hardware and computing capacity, as well as being *funders* of their own *developer* teams. These roles are especially important given the need for ongoing research and development.

7.6.2 Decision Rules and Incentives

Collective decision processes regarding changes and revisions to protocols tend to closely follow established processes of open source, typically on Github, though the nature of this mix varies by who is involved in what decisions. The process typically begins with a proposal

[5]For a self-description, see https://www.python.org/psf/summary/

[6]https://www.linuxfoundation.org/. For the Linux grant program, see https://www.linuxfoundation.org/about/diversity-inclusiveness/infrastructure/. For Python: https://www.python.org/psf/grants/.

submitted by developers following a project-specific template.[7] One of the most common trajectories for this process is to have the Core Developers vet proposed changes to the protocol and lead periods of public deliberation about those changes, but leave final decisions about implementation to provisioners who run the nodes that comprise the network. This format is often called *off-chain governance* given its lack of automation. More recent innovations attempt to bring more of the decision process *on-chain* and into the functioning of the protocol itself, though the broad outlines of the proposal-deliberation-decision process remain similar.

Following Bitcoin's broad outlines, Ethereum's protocol governance process is triggered by proposed changes to the protocol. The process of submitting an Ethereum Improvement Proposal or "EIP" (defined, as is typical in OS improvement proposals, in the first EIP published to Github, or EIP-1) is closely specified in terms of the technical and editorial format of the document, and accords significant discretion to the six named EIPs editors. These editors are expected to operate according to rough consensus in Github discussion based on strictly technical criteria. Once an EIP is approved for broader discussion, it passes into the deliberation and (for Ethereum) the second decision stage. This deliberation stage occurs over a proliferating range of websites and social media accounts hosted by the Ethereum Foundation, related groups, and leading individuals in the developer community.

In contrast to the openness of deliberation, decision-making is centralized in a small group organized by the Foundation, with most decisions made at biweekly Core Devs meeting. There are some coders who identify themselves online as either present or former Core Devs, but the exact membership of the group is impossible to know from the outside. Moreover, attendance at these meetings typically includes membership from various teams working on important clients (e.g. GETH and others sponsored by Parity and Consensys) and occasionally members of the research team, but appears to vary weekly according to the assessment of the Foundation. On one level, these meetings are extremely transparent - an agenda is published in advance, the meetings are livestreamed on Youtube, and the meeting notes are posted together with video links and

[7] The Python community established this precedent in 2000 with the "PEP" or Python Enhancement Proposal" process. See https://www.python.org/dev/peps/

agendas on a dedicated project management repository in the Ethereum organization's Github.[8] However, all discussions and decisions proceed according to rough consensus among those in the room, and there is no formal mechanism for external views to manifest in the Core Devs meeting.

Once this group has reached a decision to change the protocol, the upgrade is encoded into a new, revised version and a date is established for the transition to the new code. The intensity of the change determines the nature of the transition, with the latter described using variations on the open source concept of forking code. Those that are optional for nodes to adopt are generally described as *soft forks*, while those that node operators must adopt to remain in line with the rest of the network are called *hard forks*. The decision of whether or not to adopt the new version of the code and protocol is left to each node *provisioner*, on the expectation that they will have a strong incentive to remain in line with their peers in order to remain attached to the primary version of the blockchain rather than be stranded on an unprofitable branch. In practice, this decision can also be complicated by disagreements over protocol changes that lack any other mechanism for expression. In the extreme, this can lead to project-level forks, as I discuss below.

Approaches to bringing governance "on-chain" typically do so by assigning governance rights to the same node operators delegated block production rights. This is typically done through staking, or the commitment of project-based tokens in order to gain participation and decision rights. Stake-based voting is widely used for simple voting procedures in which votes are weighted by the number of tokens the voter is willing to bond or stake, on the general assumption that this investment creates a strong incentive for informed participation in the improvement of a project and its protocol.[9]

For example, the Tezos blockchain-layer protocol was designed to avoid the pitfalls of offchain governance by incorporating the entire process of developing, vetting and deciding on protocol amendments into the protocol itself (Goodman 2014). The proposal-deliberation-decision process unfolds across four stages, each of which is enacted by the

[8] https://github.com/ethereum/pm

[9] Another common structure is to weight votes by the product of the amount of staked tokens and the duration of time for which they are staked.

same delegated node providers or “bakers” who participate in the consensus process of producing new blocks. Proposals to amend the protocol can only come from bakers and are required to be in implementable code. If a quorum of 5% of votes cast by other bakers support the proposal, it moves onto the next “exploration vote” stage in which it must meet a higher quorum of support from bakers. Proposals which survive beyond this stage are implemented on a virtual network for 48 hours as a simulation of their impact on the actual network. This testing period is intended to provide the basis for deliberation by both the bakers and the larger Tezos community of stakeholders, though the mechanisms for this deliberation are not specified. The fourth and final stage of the process is a vote by the bakers, with the same quorum requirements as in the second stage (Arluck 2018; Goodman 2014).

7.6.3 Community Consensus

At the broadest level, an entire project may fork or split into subsets. In a project-level fork, the departing group copies the code and creates a new ledger that shares the older ledger’s history up to the point of the new ledger’s inception. Doing so allows the departing group to enact whatever version of technology and governance they preferred but couldn’t effect in the existing protocol. In such cases, the projects share a single history (memorialized in a single blockchain) up to the point of the fork - typically dated in terms of a specific block’s formation - after which they become independent chains with independent coins, separate groups of developers, separate listings on coin exchanges, etc. Those who own tokens on the original chain are typically offered an additional, equivalent number of separate tokens on the new fork, both of which at least initially can lose value because of the split.

The open nature of blockchain code and protocols means that such forks can occur for many reasons, with the primary driver being a failure to attain social consensus about a protocol’s future development. This lack of agreement can play out in one of two ways. First, the deadline of a *hard fork* for a protocol change by developers can trigger a *contentious fork* in which a part of the community disagrees with the change and secedes to create their own version of the protocol.

Following such splits, both the incumbent project and the dissident project move forward with their own resources, which are typically each a subset of those of the incumbent project prior to the contentious fork.

Because traditional protocols using off-chain governance require the building of social consensus around proposed changes, they are acutely vulnerable to situations in which there is no mechanism for arriving at or measuring that consensus. As a result, there are risks that decisions taken about changes to protocols will reflect only a subset of consensus. Success thus requires some mechanism for generating legitimacy for both the decisions taken as well as the process leading to those decisions. In turn, this means ensuring that the notion of social consensus has an *operational* pathway - it is not simply an outcome but a measurement process.

When consensus fails, it often results from a mismatch Perhaps the best-known example of this kind of failure is the case of the DAO and Ethereum. The DAO was an experiment in decentralization whose designers sought to create a democratic organizational structure that operated primarily through a combination of operational rules encoded into smart contracts and collective decision-making by members.[10]. Although the DAO was technically created by a separate organization, it was deeply enmeshed with the leadership of the Ethereum developer community. In particular, the DAO was marketed as having a set of Curators who were presented as providing "failsafe protection" under a bold heading of "Protecting the DAO."[11] Nine of the 11 named curators were high-profile figures from the Ethereum Foundation, while the other two were from Parity. Moreover, one of the co-founders of Slock.It was a core developer of Ethereum, and another was the CCO of the Ethereum Foundation.[12]

Then the DAO was hacked. This set the stage for a contentious discussion that hinged on whether the original code, however flawed, was a binding agreement or was open to change. On one side of the debate were those who demanded that Ethereum maintain the immutability of its blockchain, arguing that "code is law". The other side claimed that the scale of the hack posed an

[10]For a more thorough history of the DAO, see DuPont 2018

[11]http://web.archive.org/web/20160501142419/https://daohub.org/curator.html

[12]https://blog.slock.it/former-ethereum-cco-stephan-tual-joins-slock-it-team-9fd956f2408

existential threat to Ethereum, and that the hard fork was the best way to move forward.

In spite of the enormous stakes for the larger Ethereum project, comments by leading figures from Ethereum and Parity (all named DAO curators) in the weeks leading up to the hard fork struggled to articulate a process for decision-making in the absence of formal mechanisms. Although the terms "social consensus" and "community" were (and remain) widely used as the ultimate decision-makers, there was no definition of who was to participate or how. For example a blog post by a leading figure in Ethereum identified four primary groups in addition to developers as being engaged in discussion: app developers, exchanges, miners and "users," though the latter term (as was often the case in these discussions) was never specified, nor was a forum for this discussion mentioned.[13]

The lack of a mechanism for all of the involved roles to provide input, let alone to know whether and how their input was considered, led to a powerful backlash that resulted in a forking of the chain and the larger communities around it. The largest part of the communities remained with the chain that followed the Foundation's recommendations, which also retained the Ethereum name. A separate group split off to create Ethereum Classic, claiming that their refusal to change the ledger represented the true spirit of the project. Both projects remain in operation at the time of this writing.

Zcash provides a more effective example of this process in action. To develop and launch the project in 2016, its founders initially established a private company (the Zcash Corporation, later named the Electric Cash Corporation, or ECC) that employed the core developer team, funded in part by early investors granted equity in the Corporation. Compensation for the early developers, investors and, later, an independent foundation (the Zcash Foundation) was encoded into the protocol's monetary policy in the form of a Founders Reward, or 20% of the coins minted to pay miners for each block. The project constrained the Founders Reward to the first four-year period of mining, after which all mining rewards would go directly to miners.[14]

The protocol's monetary policy places the end of the first four-year period in October of

[13]https://www.parity.io/how-we-find-common-ground-and-settle-our-differences/
[14]https://electriccoin.co/blog/funding/

2020, creating uncertainty around the future of development that began in mid-2019. Both the ECC and the director of the board of the Foundation took positions in support of finding another funding mechanism to support further development, though the issue of whether and how to do so was hotly debated in the larger Zcash community. The Foundation's response was to develop a multi-stage process that augmented its existing ZIPs proposal structure, open deliberation and Foundation-led decision making with steps to ensure broader participation and a clearer expression of various communities' preferences.[15]

The first stage was to issue a call for proposals for how and whether to manage funding, to be submitted as ZIPs until the end of August 2019. After some negotiations between ECC and the Foundation over the ZCash trademark, the Foundation launched a poll on the Zcash forum where those who had been members for more than six months could vote on each of the 13 resulting proposals.[16] The community voted to approve two of the 13 proposals. The community vote ended in time to provide input to a more definitive vote by a Community Advisory Board of expert members, who cast their votes on the 13 proposals using a blockchain-based voting service called Helios. Miners were invited to vote but none did. The resulting votes echoed the community's feedback with the exception of approval of a third proposal, bringing the number of viable possibilities to be considered by the Foundation to three.

The three each proposed some form of continuing to use newly minted tokens created during mining/block creation to fund ongoing development by both the ECC and the Foundation, with the chosen option also incorporating a third stream of funding for external teams. After revising the ZIP into a new version with these changes, the Foundation did one last poll of community members, who voted for the overall ZIP with a clear majority but were ambivalent about how those external team grants should be governed.[17] The Foundation's decision, articulated in detail in a blog post that referenced a lengthy forum discussion, was to move forward with the ZIP with some minor modifications.[18] The final touches were publicly negotiated between the Foundation

[15] https://www.zfnd.org/blog/dev-fund-guidance-and-timeline/
[16] https://forum.zcashcommunity.com/t/community-sentiment-collection-poll-on-nu4-dev-fund-zips/35439
[17] https://vote.heliosvoting.org/helios/elections/43b9bec8-39a1-11ea-914c-b6e34ffa859a/view
[18] https://www.zfnd.org/blog/zip-1014-poll-results/

and ECC in a dialogue attached to the new proposed code in Github and were agreed on February 26th, 2020.[19]

The successful resolution of this update by Zcash was not agreed to by the entire community, however. A small group of developers and stakeholders disagreed with some of the changes and split off to form their own clone of Zcash called Ycash. While such forks are often treated as existential threats (perhaps because scholarship on open source approaches it from the perspective of large incumbent projects), this fork was viewed as an amicable separation within the community, and was termed a "friendly fork" on the central Zcash community forum.[20] The dual potential of forking as simultaneously a risk and a generative possibility gets to the heart of the deepest and most entrenched (and thus, most consequential) aspects of blockchain governance, which is its intentional incorporation of instability.

7.7 Level III: Constitutional Transformations

The site of greatest novelty in blockchain is also the dimension of governance that sits least comfortably within traditional governance frameworks. In Ostrom's framework, the constitutional level of decision-making refers to decisions about changes to the framework governing collective choice decisions. The potentially transformative effects of such decisions lead to their being deeply entrenched (Starr 2019) in structures such as written constitutions and other forms of rules that are resistant to change under any other than extraordinary circumstances. The ecology of blockchain projects has generally inverted such expectations, in effect entrenching the potential for change and instability at the constitutional level of decision-making.

The first of these is inherent to the use of open source code and public protocols, namely the ability to fork an entire project. In contrast to the simple copying of code (a different kind of fork), a project-level fork entails both the copying of code as well as the use of that code to create

[19] https://github.com/zcash/zips/pull/324

[20] https://web.archive.org/web/20201107230244/https://forum.zcashcommunity.com/t/announcing-ycash-the-first-friendly-fork-of-the-zcash-blockchain/33162

a new version of an existing project. Although the presence of a foundation or another entry with rights to the original name means that only the first branch will retain the original name, there is often nothing to keep those creating the new fork from using the existing code under a new name. Such forks are always a risk in moments of contentious collective choice decisions for open source projects, during which the potential for part of the network to secede and create their own version is an ongoing threat. The centrality of open source to blockchain as a technology thus embeds what theorist Unger terms “destabilization rights” as implicit affordances to actors beyond those charged with collective choice decisions (Unger 1987). While carrying risks, these destabilization rights may open up more radical alternatives in early-stage technological development than would otherwise emerge.

The second driver is a general commitment to decentralization of governance as an aspirational process rather than an immediately attainable outcome. The field-level commitment to what practitioners term “progressive decentralization” is reflected in widespread experimentation with new modes of technologically coordinated collective action that can then be selectively introduced into existing projects, with successful experiments providing templates for the diffusion of those modes across other projects. These experiments are occurring in both blockchain-layer and blockchain-enabled protocols but are most prevalent in the latter given the flexibility and affordances of smart contracts as encoded mechanisms. These experiments are often combined with fully encoded or distributed autonomous organizations, which are successors to the original DAO that reflect the lessons learned in that project’s failure.

7.7.1 Changes in Roles and Rights

The open source nature of protocols, together with changes in technology, means that both protocols and the larger projects they define are continuously refined and revised. Much of this refinement is the result of the ongoing development of new forms of smart contracts. These mechanisms combine the potential of code-based money to constitute new systems of coordination (Desan 2014; Halaburda and Sarvary 2016; Pistor 2019) with the ongoing advances in mechanism and market design described above. This combination of methods, often gathered under the neologisms of *cryptoeconomics* or *token engineering* join financial market mechanisms with cryptographic and game theoretic methods that allow for much more finely grained and targeted incentives and penalties for each of the roles specified in the protocol (Cong, Li, and Wang 2019a).

Projects also often take the same approach to the governance of their protocols by treating it as an ongoing, adaptive process that shifts as the range and salience of stakeholder groups change, and as new modes of collective governance become available. At the level of the ecology, this approach opens each of the four stages of proposal, deliberation, decision and implementation to greater and greater decentralization, meaning greater control by stakeholders (see e.g. Fu 2020). Although not universally, this change often involves the introduction of staked voting mechanisms into projects once they have become established. Sometimes this involves the creation of a new token for an existing project that confers governance rights in that project, while in others the same token is used for all purposes within a project in order to link incentives more tightly.

0x is an example of the latter. After a first year during which the project's protocol was under the sole control of its core developers, the project's cofounder in 2018 wrote a blog post summarizing the project's research into various governance structures, as well as its long-term plan to shift control of the protocol from the founders (via the company they founded) to the larger community of stakeholders using and building on the 0x protocol (Warren 2018). That vision included introducing staked voting using the project's existing ZRX token as well as a series of projected milestones intended to mark progress toward broader stakeholder governance. The project updated this roadmap with another blog post in early 2020 that announced several new projected milestones and new plans that included a reworking of the "ZEIP" or ZeroEx Improvement Proposals to potentially include an expanded set of roles, action rights and decision rules to govern the process of proposing, deliberating, deciding and executing changes to the protocol (Gonella 2020). These changes would expand the named roles from their original three – makers (of liquidity), takers (of liquidity) and relays – to include others including operators of mesh nodes, market makers and stakers. These latter roles are all related to expanding and deepening the liquidity the overall network is able to offer.

7.7.2 Incentivized Social Consensus

More broadly, the sense of decentralization as a processual goal rather than an immediate binary has also resulted in experimentation with a wide range of possible modes of governance. Some of these modes emphasize broadening participation in one or two components of the proposal-deliberation-decision-execution tetrad, while others apply all four stages to develop digital alternatives to established governance institutions. Their primary points of commonality are the use of smart contract code to enable collective forms of problem identification and solving, as well as their modular development and use across multiple projects.

In addition to the implicit curation provided by reviews and reputation, many projects incorporate a more formal curation structure known as a Token Curated Registry, or TCR (Goldin 2017). A TCR is a collectively generated list that draws its legitimacy from its ability to use tokens to incentivize its community to provide useful information about the list's components and to vet members' contributions. The basic idea is that candidates will be motivated to propose their own inclusion into a well-curated and relevant list (e.g. DistrictOx's DNT token and set of autonomous component markets, or AdChain's AdToken and white list of approved advertisers, or Messari's list of researched tokens), while the owners of the tokens attached to that list will be motivated to maintain its high quality. Token owners have the option of acting like passive owners, in which case they do little other than invest by buying the tokens and hoping for a price gain, or becoming actively engaged in the evaluation of new proposals, in which case they theoretically benefit by improving the information value of the registry and thus increasing the token price still further.

In addition to predicting prices and events, prediction markets also provide a mechanism for collective decision-making through implementations of "futarchy" (Hanson 2013). In futarchy, the process of voting is restricted to establishing a societal consensus about big-picture priorities.

The uncertainty about which policies to use to reach those goals is addressed by policy-specific prediction markets, with the policy predicted as being most effective being the one chosen. This distinction between "voting on values" from "bet[ting] on beliefs" was later modified by Merkle (Merkle 2016), whose model replaced voting altogether in favor of market-based predictions.

Quadratic voting (QV) is a voting method that attempts to lessen the tendency of economic and political systems toward concentration or even plutocracy. Developed by Glen Weyl and a series of coauthors, quadratic voting is part of a larger set of policy recommendations under the banner of "Radical Markets" (Lalley and Weyl 2018; Posner and Weyl 2018). QV assumes some form of weighting for each individual voter based on social influence, wealth or some other unequally distributed characteristic. Individuals' votes are counted based on the square root of their weighting, bringing the votes of those with the highest weights closer in line with those at the bottom of the distribution and thus magnifying the relative impact of the latter's votes relative to other weighting schemes. Another is delegative or *liquid democracy*, an approach first advanced by computer scientist Bryan Ford (Ford 2002), with a dual objective of increasing democratic participation while also ensuring that votes are cast by those with the greatest commitment and expertise. Liquid democracy does this by giving individuals the option of voting themselves or delegating their votes to others, who may themselves delegate votes to specialists. Various forms of both kinds of writing are widely used in the governance of both blockchain protocols and the larger projects they support, and are particularly prominent in DAO-like structures.

7.7.3 The Rise of DAOs

The modular affordances of encoded governance are perhaps clearest in the increasing use of distributed forms of organization in which interaction between members is largely mediated by collections of smart contracts. Such distributed autonomous organizations or DAOs typically incorporate a mixture of their own, customized rules as well as one or more collective governance components from the list described above (i.e. curation, prediction markets and decision-making). DAOs are themselves platform-based modules that can be incorporated into particular governance processes, while also allowing for the selective incorporation of the mechanisms described above.

Blockchain-enabled platforms such as Aragon or DAOstack that enable the launch of DAOs typically incorporate these mechanisms as options for the creators of DAOs to combine and reshape as best suits the nature of their project. These DAOs include both independent entities as well as the governance of existing blockchain projects. For example, the dxDAO - a token exchange created by prediction market Gnosis (one of the most prominent blockchain-enabled protocols building on Ethereum) - draws on a similar set of tokenized mechanisms, though it combines them differently than does Cement. The dxDAO runs on the DAOstack protocol, and also makes use of DAOstack's "holographic" consensus algorithm for its own governance. Holographic consensus proceeds from the assumption that attention is the scarcest resource. To manage this, the protocol separates voting (and thus voters' attention) from economic power, with the latter exercised through prediction markets used to "boost" proposals to voters' attention and also to signal market actors' views. In keeping with the general pattern of prediction markets, all parties are rewarded (penalized) for their alignment (lack of alignment) with the majority view, which is taken to be the consensus.

The modular nature of smart contract-based governance also allows for the development and recombination of elements by individual projects that incorporate DAOs without relying on platforms such as DAOstack or Aragon. For example, Cement is a stablecoin (a token designed to serve as a reliable store of value by maintaining a relatively stable price) that has from inception been designed to be governed by a DAO, the CementDAO. In addition to its stablecoin, the project incorporates a distinct governance token (in addition to its MIX stablecoin) as well as two primary forms of staked decision-making that each confer participation rights in various aspects of the system's governance, primarily involving the composition and relative weightings of the components in the basket of other stablecoins in which Cement invests in order to support its stable price. DAO members collectively choose delegates who have in turn staked and locked the project's governance token, the BILD, in order to participate in the curation of a list of approved stablecoins in which to invest. Decisions about the construction of the portfolio are in turn made through a futarchy-based prediction market in which decisions are based on the market value of BILD (in MIX) established in paired markets, one for the proposal and one against it.

DAOs have also become a structure used to match developers and funders working across multiple projects. Many of these have formed by copying the code of a project called Moloch Dao, which was built with the dual purpose of pooling funding to be distributed based on member votes for the development of new projects and functionality on Ethereum, as well as providing a forkable codebase for other groups interested in organizing their own collective organizations using its template. Such groups have the potential to provide an entirely new source of funding beyond that provided by centralized organizations and foundations, though at the time of this writing their funding streams are small relative to those mobilized by the established project foundations.

7.8 Discussion: Open Source Institution Building

The multilevel, multi-variable Ostrom framework I develop in this chapter provides several insights into the innovations and challenges involved in protocol-based governance in blockchain. I summarize my findings in the context of that framework in Figure 7.3, below.

Figure 7.3: Summary of Findings

	Roles, Rights	Incentives, Rewards	Decision Rules	Shared Expectations
Operational Governance by Protocols	Participation rights	Algorithmic/ encoded	Algorithmic/ encoded	Implicit
Collective Choice Governance of Protocols	Decision rights	Algorithmic Financial Status	Institutionalized OS mechanisms	Explicit
Constitutional Governance Transformations	Destabilization rights	Contingent, redistributed	Progressive incorporation	Divergent (forks), convergent (DAOs)

Figure 7.3 points to a resolution to the seeming paradox of change and control that characterizes protocol-centric governance in blockchain. The table's levels reflect the dual

hierarchy I described in the framework in Figure 7.1, with a bottom-up *hierarchy of control* and a top-down *hierarchy of change*. Success at the operational level is predicated on the ability of the protocol's encoded roles, rights, incentives and decision rules to encompass the interests of the projects' designers, participants and other stakeholders. When that is no longer the case, a project will be viable only if it can successfully navigate the collective choice process of revising the protocol. Rigidities at the collective choice level means that profound change often happens at the field level, through the creation of new projects. The expansion of projects built on blockchains (as opposed to being blockchains themselves) has enabled widespread experimentation in individual projects' protocols that then go on to become more widely adopted building blocks of larger efforts. This is due to both the normative impetus toward decentralization and the increasing sophistication of smart contracts, both in their own right and as a means of combinatorial innovation (meaning, concatenating the smart contracts to create new collections of them)

The most unusual aspects of governance in blockchain stem from the sociomaterial aspect of interwoven changes in both the technology and the governance practices that shape its use. These intersect most directly in the use of open source techniques for both technology development and for decision-making, giving each of the three levels of governance in blockchain a shape that is distinctive from that of traditional forms of governance and social organization, while also inheriting key characteristics of open source. Put another way, the institutionalization of open source as a development method for software has in blockchain resulted in the incorporation of open source methods in the co-production of both organization and technology. This open source institution-building is a natively digital alternative to traditional means of building governing institutions, just as the inside-out organizational structures I described in the prior chapter are natively digital means of coordinating resources. I explore the implications of these structures in the chapter that follows.

Chapter 8

Conclusion

The nature of conferences changed drastically in the final year of my project as COVID-19 swept the world. The last conference I attended was notable in part because I was able to watch real-time presentations and interactions between leading central bankers from my sunny home office on a spring day in 2021.[1] The premise for the conference was a showcase of innovation by central banks, and opened with a session moderated by financial journalist (and trained anthropologist) Gillian Tett of the Financial Times. Tett deftly led the panelists through a series of questions about the blockchain-based innovations in digital financial infrastructure and money by the institutions they led, which included the Federal Reserve, the European Central Bank and the Bank for International Sentiments. She closed the session with a quote from a recent popular business book by IBM's former CEO titled "Who Says Elephants Can't Dance?" about his restructuring of the technology giant (Gerstner 2002), and noted how central banks were clearly already "dancing" with this new technology.

Tett's gambit recast the vaguely comical scene of investment bankers doing the chicken dance from the outset of my project in a new (if more intentionally stately) light that reflected both the importance of blockchain as an infrastructural technology, and the ongoing nature of experimentation involved in its implementation. This combination of serious investment and

[1] Archived conference page available at https://web.archive.org/web/20210610125219/https://www.bis.org/events/bis_innovation_summit_2021/overview.htm.

continued “ferment” (Tushman and Rosenkopf 1992) as the technology remains unsettled has characterized the entire ecology for the duration of this project.

While this period of experimentation is ongoing as I conclude in 2021, my research provides insights into the structure and processes that characterize the multi-level, sociotechnical organization of this early stage of digital transformation and the role of blockchain technology within it. The analysis of Chapters 3 through 5 first builds a methodological approach to de novo classification in the absence of an authoritative taxonomy, then applies that approach to develop a sociomaterial analysis of the technology as a GPT by addressing the dual challenge of identifying such a technology: identifying the new technology’s components and their generality of purpose, as well as identifying their patterns of diffusion across existing domains of application. The three unique aspects of the technology’s generality of purpose, I find, are its ability to create generative digital/computational infrastructure, the creation of programmable money for payments, and the ability to create and manage digital representations of both physical and virtual objects, and to link those representations to incentives to which they give force. I also find a fourth general pattern in my data defining a form of organization common to the implementations I study. These findings motivate the closer, case-based study I undertake in Chapter 6.

Although I discuss the carriers of the technology’s generality of purpose in isolation in the analysis that comprises chapter 4, their very generality means that they are most often used in combinations of one or more of the components. The taxonomy and classification system I develop in Chapter 3 also underwrites my ability to explore patterns of variation in these combinations of components across the larger landscape of implementations captured by my unique data set. My findings emphasize the interplay of institutions and the technology, and in particular the drivers of variation in the re-invention of the technology as the core mechanism of its diffusion (Rogers 1995). This variation, I find, is a function of the specific ways in which the carriers of the technology’s general purpose nature are used and combined in the context of the salient institutions in a given application domain. More specifically, I find that the nature of reinvention varies with the extent of overlap between the technology’s function in a given domain and the function of existing core institutions. I trace this variation across three patterns of diffusion, each of which is defined by a specific form of digital reinvention. The patterns I identify as speculative and competitive reinvention are marked by significant alignment between the core institutions (e.g. money, law) and the functions of the technology (e.g. cryptocurrency or tokens, smart contracts) that entrepreneurs and developers are exploring. The primary differentiator between these types of reinvention is the presence or absence of incumbent industries and institutions (e.g. central banks) in developing their own digital alternatives. The contrasting pattern – pragmatist reinvention – is the realm of "use cases" in existing industries seeking to solve well-established problems while retaining much of their fundamental structures.

My findings at the meso level of organization bring a relatively new institution into the discussion: open source processes. I find that the institutionalization of open source as a mode of code development, along with the pivotal role of developers in blockchain technology, has resulted in new modes of coordination and control that upend existing theories. Coordination in blockchain implementations builds on individual components (e.g. platforms, ecosystems and communities) that are individually subjects of active research and theory development, but - like open source itself - have not yet been studied as components of larger sociotechnical assemblages. I find that forms (particularly the mediating hierarchy) and practices originally developed for the coordination of open source have migrated to become central to the coordination of blockchain in ways that build more fluid and negotiated ensembles.

My findings at the meso level of organization bring a relatively new institution into the discussion: open source processes. I find that the institutionalization of open source as a mode of code development, along with the pivotal role of developers in blockchain technology, has resulted in new modes of coordination and control that upend existing theories. Coordination in blockchain implementations builds on individual components (e.g. platforms, ecosystems and communities) that are individually subjects of active research and theory development, but - like open source itself - have not yet been studied as components of larger sociotechnical assemblages. I find that forms (particularly the mediating hierarchy) and practices originally developed for the

coordination of open source have migrated to become central to the coordination of blockchain in ways that build more fluid and negotiated ensembles. More specifically, I find that projects implementing the technology leverage the affordances of tokenization in order to create digital assets that enable the management of complementarities between the interests of the various communities involved in co-creating the resources necessary for the implementations to function. I also find that the very fluidity and loose coupling that make these organizational configurations difficult to map to more traditional (and, theoretically, more durable) organizational structures are the sources of their persistence in an environment of constant technological, regulatory and market change.

The influence of open source is even more evident in the governance of these implementations, which reflect the technology's origins in cypherpunk and other efforts to create computational and cryptographic alternatives to core institutions such as money, property and law (Brunton 2019; Swartz 2017). What differentiates these newer variants is their ability to build on these earlier, foundational efforts and extend the technology's reach into more operational aspects of governance such as organizational coordination and market design. They also differ in the breadth of the communities they seek to mobilize and incorporate through open source methods. Where the foundational work was largely theoretical and was discussed in private listservs, this new work is open in every sense: to outside scrutiny, to review of the code, to broad participation through tokens and through the forking of codebases for experimentation by others.

This *open source institution-building* inherits both the institutional objectives of the cypherpunks as well as the mores of open source development but applies them in a wider variety of settings. Its closest institutional analog is the pragmatism of experimental governance (Ansell 2011), in that both create spaces for experimentation in the context of larger institutional infrastructures. That similarity is quite weak, however, given that the potential for transformation in the underlying infrastructures in blockchain technology is built through open source destabilization rights, while traditional state institutions are typically built with the opposite imperative (Starr 2019).

Open source institutions have the potential to significantly reshape much older, incumbent institutions by creating sufficiently compelling alternatives that require an institutional response. For example, the rise of stablecoins in 2020 and 2021 has been driven in part by the ability of new projects to fork influential protocols in order to rapidly build their own variations, such as the Curve protocol for stablecoins. While it is impossible to measure the extent of this effect, it is clear in communications from central banks and related institutions (primarily the BIS) that the rise of stablecoins - and not Bitcoin - is now seen as a primary impetus to the development of central bank alternatives.

The open source institution building in blockchain also has the potential to interact with (and thus shape) the trajectories of other GPTs that are part of digital transformation. By offering core institutional infrastructure such as property rights and payments, as well as more complex applications such as marketplaces for big data (whether from corporations or autonomous networks of Internet of Things sensors) and artificial intelligence/machine learning models, blockchain technology is already heavily engaged with its emerging peers. This imbrication hints at one possible trajectory for digital transformation in which natively digital institutions matter at least as much to the governance of new technologies as any traditional institution.

8.0.1 Implications for Future Work

A research project studying an emerging technology is more of an ongoing commitment than a neatly time-bound thing. The contributions of this dissertation thus remain affixed to a specific technology and point in time, though they provide some indicative findings about the trajectory of digital transformation in the coming years, and the several ways in which sociology can contribute to our understanding of this transformation as it is happening.

First, my work points to the potential fruits of treating classification and taxonomy building as an analytical tool for the study of *emergent* macrosocial organization, while building on the rich insights and literature within and outside of sociology on processes of classification and categorization. Bringing these insights into methods development, and in turn bringing machine learning more directly into dialogue with coding methods from qualitative sociology, have allowed me to develop not only novel methods but also a potentially toolkit for developing insights into large-scale technological change in medias res.

This project also clarifies the need for organizational and economic sociology to engage with platforms, ecosystems and open source communities as explicitly organizational structures, as a complement to already important work happening around the implications of these structures for the economic well-being of their participants. Although the pioneering work on organizational networks as a form in the 1990s (Powell 1990) laid the groundwork for the expansion of work on these organizational forms in the economics and strategy literatures, each has the potential to provide insights that expand the scope of 20th century organizational theory beyond its emphasis on the Weberian bureaucracy as the explicit or implicit unit of analysis.

My project points to two possible ways such an expansion might be effected. The first is through more consistent attention to "organization" as a sociomaterial phenomenon, a fundamental approach taken in the literature on technology and organizations that is essential to broader analysis of technological change. While this approach requires a deeper engagement with *both* the potentially complex and technology and the potentially new forms of organization imbricated with it, the potential analytical rewards are accordingly more compelling. It also offers the potential to recenter organizational design as an agentic concern for organizational theory that brings the focus back to the meso level of social organization and behavior.

The second, and closely related, possible pathway is through a shift away from a search for "the" definitive form, and toward an embrace of both the heterogeneity and fluidity that characterize sociotechnical organization today. Ironically, I have found in this project that the former star actor - the formal bureaucracy - provides a compelling point of entry for analysis of these forms, though this role is only clarified by avoiding easy binaries about either the "end" or the dominance of the corporation.

More broadly, my work extends on recent efforts in economic sociology to incorporate the study of blockchain and its component technologies into broader conversations. While economic sociologists were among the first in the discipline to appreciate the implications of cryptocurrency in the context of a rich tradition of studies of the nature of money (Bandelj, Wherry, and Zelizer 2017; Dodd 2014, 2017), there has until recently been relatively little attention paid to the other aspects of the technology that I explore in this work (but see Caliskan 2020). The general-purpose aspects of the technology and the nature of their diffusion through the creation of digital assets, market infrastructure, money and other applications point to rich potential for the continued extension of economic sociology's insights into 20th century versions of these institutions into the digital era.

Finally, this project points to the importance of bringing the institutionalism back to the fore in the study of technological change in the present. The analysis of power in the digital era necessarily requires a study of both the construction and governance of algorithms and other means of control. The power of these technologies is largely conditioned on the nature of the existing institutions whose functions they assume. As with organization, a hybrid sociomaterial (or, in this case, techno-institutional) that engages with both the technology and the institutions it complicates is necessary to grasp the nature of the transformations underway.

Appendix A

Appendix: Notes on Studying the Present

I noted in the introduction to this dissertation that there are no methodological guidelines for a study of the present. This is not to say that there were no benefits – the opportunity to study a process like this in medias res comes along rarely, and if nothing else allowed me a chance to avoid the problems of distant knowledge inherent in historical studies of technology-in-use (Barley 1990). Moreover, the institutionalization of open source as an operating model, the shift to multistakeholder communities and a general grounding in the knowledge commons of open science (particularly for scholarship in cryptography, systems design and law) meant that the amount of potential data far exceeded what I could ever possibly use.

As a result, any single ethnography, quantitative study, participant observation, or other focused approach might provide deep knowledge of a single site or data set, but would almost certainly prove to be too narrow to address the more generalized questions I study. That's why this is a mixed-methods dissertation of necessity. A commitment to following the actors means identifying and spending time in the spaces where the work of developing and implementing the technology is happening. I do this by focusing on three loosely defined sites, each with its own set of data and its own methods. My methods ultimately involved triangulation across all three. For the sake of clarity, however, I present them here individually, beginning with the physical spaces in which community members interact.

A.1 Fieldwork and Interviews

I began my research by attending a series of industry conferences, Meetups in New York City and Philadelphia and closed working groups through the Coalition for Automating Legal Agreements, or COALA. These interactions were foundational to developing an understanding of the overall ecology of actors involved in the technology, as well as the dynamics of status and power between communities of developers, investors, project founders and the broader public. This dynamic was particularly fraught when I began my field research in 2017 amidst a year-long boom in token prices and initial coin offerings or "ICOs". This boom was intrinsic to (and perhaps a partial result of) the surge in public interest I noted above, and was in retrospect an unusual opening for me as a researcher given the influx of others also seeking to learn about the technology.

Where my reading on the topic was focused on building knowledge, my emphasis in interacting in these spaces was on attempting to disambiguate some of the mystique around the technology and its uses in practice. This approach reflected Karl Weick's description of new technology as "equivoque," which he defines as "something that admits of several possible or plausible interpretations and therefore can be esoteric, subject to misunderstandings, uncertain, complex, and recondite" (Weick 1990, p. 1).

Weick's insight is strikingly relevant to blockchain technology, which is sufficiently complex and "recondite" that only a small percentage of those involved with it (in my assessment) have a firm grasp on how it works. At the same time, and perhaps because of this complexity, there were (and remain) a series of terms that are constantly invoked as definitional of the technology: decentralization (typically referring to structures in which powerful central actors such as Facebook are disintermediated), trustlessness (referring to the possibility of action without trust institutions) and immutability (referring to the tamper-proof nature of data stored on blockchain ledgers). In my fieldwork, I encountered wildly contradictory approaches to these terms. When pressed to define them, as I often asked interviewees to do frequently in my field work, non-technologists often found it difficult to land on a single definition of what these terms meant and how to determine their validity. A conversation I had during a lunch break at one of the 2017 conferences is illustrative. My interlocutor was a senior, non-computer science academic from Europe who had come over for the week of conferences, and was a very enthusiastic believer in the technology, explaining when I asked why by referring to the threefold mantra (decentralization, trustlessness and immutability) as though these terms were sufficient explanations in themselves. When I gently pressed on the definition of decentralization, and asked how a technology that only a small minority could program could be fully decentralized, he became flustered, told me to "read Satoshi's paper" (the original Bitcoin white paper, a classic argument from authority move in these conversations) and excused himself.

By contrast, technologists are among the most skeptical of these claims, and often mobilize claims of failure along any of these dimensions as the basis of their critiques of specific protocols and projects. This skepticism was strikingly clear during a closed-door workshop hosted by Princeton and the Defense Advanced Research Agency (DARPA) I attended in October 2018 with representatives of DARPA as well as leading academic computer scientists. What was most striking about the academics' presentations was that they broadly described the construction of secure blockchains as a scientific pursuit in which each new protocol identified and addressed shortcomings in a predecessor, even though those predecessors often remained in widespread use in the field. A discrepancy between scientists and the lay public regarding a complex technology is not in itself surprising; what is striking here is the investment of the same terms (decentralization, etc.) as ideals toward which technologists are working, and the understanding of this shortcoming as a goad to further normal scientific development rather than a fundamental failure. These examples, while anecdotal, point to a much larger pattern in which a small set of concepts take on great importance even as they carry ambiguous and even contradictory meanings. Put in Weick's terms, blockchain technology is "equivoque" par excellence.

Rather than binaries that exist or fail to exist, I take two of these terms - decentralization and trustlessness - as evidence of two dimensions along which both the technology and the forms of social organization involved in implementing it vary. Doing so grounds this abstract technology in long-standing sociological discussions regarding the tensions between centralization and decentralization (perhaps better framed, as I do below, in terms of differentiation and integration) and the functioning and production of trust in exchange (Granovetter 2017; Zucker 1986).

My primary discoveries during this period were the proliferation of communities and roles around individual projects, as well as the emphasis on building these communities by projects. That said, it also quickly became apparent that I needed to push further into the substance of the technology itself if I was to understand the technical aspect of sociotechnical organization.

To do this, I extended my fieldwork with a series of 15 semi-structured interviews with practitioners active in the ecology. These interviews provided significant clarification of my understanding in that they allowed me to ask not only about individual projects but also clarifying questions about terms that appeared to be taken for granted as important but were often used in ways that implied different definitions (e.g. “what do you mean by ’governance’?”). Over time, it became clear to me that these interviews were also valuable but more as a secondary and clarifying data source for triangulation than as a primary data source.

Time was a particularly important factor here, as the rate of change in the technology rendered any insights gained from a particular interview potentially out of date within a year. For example,....

> “What is happening in this space is that new forms of governance are emerging, ones which don’t require a foundation in the sense of a [registered nonprofit] or a legal entity that is set up for the foundation, but one in which the disbursement of tokens, or of anything else for that matter, can really be governed using on-chain mechanisms, which would allow the devolution of governance away from [the founders], to I think over time, the [project’s] token holders. And I think that’s really interesting, but an area of crypto that is not yet fully developed.[1]

Two years later, DAOs had evolved to become one of the primary means of coordinating resources in new projects. This rate of change is endemic to the overall ecology, and ultimately

[1] Interview of August 8, 2018

motivated my shift away from point-in-time fieldwork and interviews to more generalizable sources of data for my questions.

A.2 Algorithmic Ethnography

My second research site complemented the first, and ultimately became more important to my project. Over the course of research, I spent considerable time in the digital spaces where much if not most of the daily work of building these implementations is coordinated, negotiated and documented (Daniels et al. 2017; Lupton 2014). For my own education, the most fruitful sources of data in this regard have been the web-based sites falling on the middle of this spectrum, which have included project-specific online discussion forums, blog posts (most of which, by convention, are hosted on Medium.com) and other sites where changes to functionality, governance issues and other topics are presented and discussed in greater depth than is feasible on social media and at a more accessible level to non-technologists than is typical on Github.

These sites were particularly fruitful for triangulating between documents, websites, fora and other online spaces where the ongoing work of building, maintaining and revising the protocols and code at the center of implementations took place (Christin 2020; Seaver 2017). These web-based sites are arrayed on a rough spectrum, with open source code development and storage site Github at the ostensibly technical end and social media sites such as Twitter at the other, though in practice discussions on both sites have both social and technical elements, with both providing infrastructure for these dispersed communities to communicate, establish identity and coordinate their activities (Kogut 2000; Orlikowski 2002).

The understanding I developed through that work (which is ongoing, befitting the rate of change in the technology) in turn contributed to the algorithmic ethnography that drives Chapter 7 (Christin 2020; Seaver 2017). Triangulating across protocol documentation as well as the sites I describe above provided invaluable data for my research into protocol governance by allowing me to trace the processes of coordination, deliberation and decision-making.

A.3 Building a New Data Set

My third research site/source of data, and the one on which I have spent the most time, is the documents related to implementations of the technology. Gathering this data initially began as an attempt to make sense of the spectrum of ways the technology was being developed and used, and to zero in one what constituted a proper unit of analysis. Even defining what to document entailed decisions that would shape the rest of my dissertation. For example, Ethereum is a major project in the ecology, and has multiple organizational units within the overall project. Should I document "Ethereum" as a single entity, or take into account all of the individual organizations, or perhaps just focus on one of the latter? Moreover, how should I approach multiple projects within Ethereum (for example, the Ethereum Enterprise Alliance consortium project)?

Following the actors and the technology led me to take individual implementations, rather than individual organizations, as my unit of analysis, and to focus on the sociotechnical organization of those implementations. So, in the case of Ethereum, I incorporated entries for Ethereum, and the Ethereum Enterprise Alliance consortium (which is developing several protocols based on but distinct from Ethereum itself) as individual units for the data set. More broadly, I have restricted my data set to projects that had at least drafted a white paper and had a website, which I treat as reflecting the intent to communicate beyond the initial project team as a prelude to gathering external resources. This boundary condition aligns with standard definitions of emergent organizations, which emphasize intent as a necessary condition (Katz and Gartner 1988). While this approach has constrained my data to a subset of all blockchain projects, collecting white papers has also provided a crucial protection against survivorship and other forms of bias inherent in studying formal data sets of new ventures (Yang and Aldrich 2012).

For each project in my data set, I document basic information including the project's name, its primary URL, the ticker of its crypto token (if relevant), its top-level Github repository, as well as other identifying information. Figure X provides a snapshot of the top of the spreadsheet in which I store the summary data. The second and third columns reflect the categories and sub-categories to which I assign each project, and imply a factual status for these groupings. In reality, the process of developing and assigning these categories was central to the development of my understanding of the overall ecology, and to the abductive development of my dissertation.

Figure A.1: Basic Database

	A	B	C	E	F	G	J	N	S	T
262	project	top level category	secondary category	url	Notes	Concept	chain	token	github_org	
263	ABChain	advertising	advertising	https://ab-chain.com/				RTB	AB-CHAIN-com	
264	ACANetwork	advertising		https://aca.network/				ACA	acanetwork	
265	Adbank	advertising	media	https://adbank.network/			E	ADB	NA	
266	AdChain	advertising	advertising	https://adtoken.com/		The adChain Regist	E	ADT	AdChain	
267	AdCircle	advertising		https://adcircle.net/				ADO	NA	
268	AdEx	advertising	media	https://www.adex.network/			E	ADX	AdExBlockchain	
269	AdFunnel	advertising		http://www.adfunnel.tech/				FNL	AdFunnelTech	
270	AdPerks	advertising	marketing	https://adperks.io/				PERK	NA	
271	AdPump	advertising	advertising	https://adpump.io/				ADP	adpump	
272	AdSigma	advertising		https://adsigma.io/				ADSi	adsigma	
273	Aref	advertising		https://ico.aref-platform.io/				AREF	NA	
274	Arround	advertising	AR	https://arround.io/				ARR	ArroundPlatform	
275	ATMChain	advertising	media	https://www.atmchain.io/		Comprehensive sc	E, own late	ATM	ATMChaiNATMChain	
276	Baner	advertising	social	https://ico.baner.su/ - documents				BNR	NA	
277	BehaviourExchange	advertising	internet service	https://behaviour.exchange/		More of a service for website		BEX	NA	
278	Benepit	advertising	phone calls	https://www.benepit.io/				BNP	NA	
279	Benjacoin	advertising	advertising	https://www.benjacoin.com/		Adding new capab	E	BENJA	NA	
280	BigBom	advertising		https://bigbom.com/				BBO	bigbomio	
281	Bitclave	advertising	advertising	https://www.bitclave.com/		Advertising and se	E	CAT	bitclave	
282	Bitcomo	advertising	advertising	https://ico.bitcomo.com/		Ad network for IC	E fork	BM	Bitcomo	
283	BlockchainProgrammaticCorporatio	advertising	algorithmic	https://bpc.one/en/				BPC	BPCPlatform	
284	Brave	advertising	advertising	https://brave.com/		Browser for mone	brave	BAT	brave	
285	c0ban.tv	advertising	with video, micropaymen	https://www.lastroots.co/	Japan			NA	c0ban	
286	CrestToken	advertising		https://www.cresttoken.com/				CSTT	NA	
287	CromHub	advertising	advertising	https://www.cromhub.com/				CROM	cromhub	
288	CryptoAds	advertising		https://tokensale.cfc.io/				CRAD	NA	
289	DANService	advertising		https://dan-service.com/				DANS	NA	
290	DATA	advertising	data exchange (unclear)	http://data.eco/				DATA	Blockchain-DATA	
291	Dysnomia	advertising	digital signage	https://dysnomia.top				MIA	NA	

Over the first year of my research, my focus within this process shifted from documenting all projects that met my criteria to one of documenting only projects for which I had one or more documents. Although I continued to collect this data for those projects (and look forward to using it in future projects), I also began to focus more on the documents themselves, and on the wealth of potential information they provide. This led to a second fundamental question: when to stop gathering new projects and new documents?

This is a deceptively simple question given the ongoing and unpredictable development of the technology, as well as its continued spread into new areas. Stopping too early would have risked leaving my data overwhelmed by whatever trends were occurring until that point, and would have rendered my attempts at generalization (let alone a comprehensive taxonomy) comfortingly but speciously easy. This is particularly the case given the wave of hundreds of new cryptocurrencies prior to 2013, and then thousands of new tokenized projects raising funds by issuing tokens through an initial coin offering or “ICO” in 2016 and 2017, when my research began. Some analysts have ignored these distinctions by referring to all cryptocurrencies and project tokens as “cryptocurrencies,” while others take the stance of labeling every token an “ICO,” neither of which are descriptively accurate as generalizations, and both of which ignore the equally important implementations of the technology in the corporate and state sectors that definitionally do not do an ICO or have a cryptocurrency.

My primary objective in this regard was to avoid having my data set become an overwhelmed by the thousands of ICOs from 2016-2017. I worked to avoid this first by going back to cryptocurrencies and protocols that were developed prior to the ICO wave, and downloading as many available project documents for those entries as possible. I also attempted to capture as much as possible during the ICO wave and after it, using (as noted above) websites specializing in presenting ICO information. Doing this was crucial given how many of these projects failed, taking their websites and documents offline in the process, and was important in order to avoid the risk of survivorship bias in my data (Yang and Aldrich 2012).

Doing this gave me an initial set to which I continually added based on coverage of new projects in industry press and newsletters, as well as on social media (primarily Twitter and LinkedIn, for which I built a profile for this purpose). The latter sources were particularly important in gathering data on new projects that did not launch a new token, especially those launched for enterprise or large corporate clients, as well as for infrastructural blockchain technologies launched by major corporates (e.g. Microsoft's Azure hosting for blockchain, or IBM's FoodChain project) and trial implementations by central banks and other state agencies around the world. I continued gathering these into early 2021 given structural changes in the ecology that include the rise of "defi" (as I describe it in chapter X) and central banks' rapid convergence on issuing digital versions of their currencies or CBDCs (Central Bank Digital Currencies), which I would have missed had I stopped in even early 2020.

A.4 Building a Corpus

The primary aspect of the project-level data I describe above is not the spreadsheet of projects, but rather the project-level documents I gathered for each project in the data set. Project white papers have been the most important source of data for my project. The norm of writing at least one white paper for a technological innovation is well established across the entire technology industry, and was also established by Bitcoin's pseudonymous founder(s?) Satoshi Nakamoto, who introduced Bitcoin with a white paper. Although not universally followed, the norm of crafting a white paper to introduce new projects is sufficiently established that nearly every project seeking to raise funds publishes at least one such paper that describes some combination of the project's technology, business objectives, mission, community strategy and other factors. These papers reflect the melding of social and technological organization in blockchain projects in that they blend the technological practice of producing technical papers introducing new technology with the entrepreneurial practice of writing and circulating business plans (Baden-Fuller and Morgan 2010; Ghaziani and Ventresca 2005). In contrast to privately circulated business plans, however, blockchain project white papers are publicly posted online, and are a central point of references for the various external communities within which projects interact. This circulation makes them among the most important boundary objects in the overall ecology (Doganova and Eyquem-Renault 2009; Star and Griesemer 1989).

Projects that produce more than one paper generally split these concerns into separate papers for technology (often called a blue, yellow or green paper) and a more commercially oriented white paper, while more purely technology-oriented projects (e.g. those functioning within large organizations, such as Microsoft's Coco Protocol) tend to publish purely technology papers. Given the early and constantly changing nature of the technology and the broader ecology, projects often update their papers steadily over time as the their technology and business models evolve. As a result, Bitcoin is both a model for later projects as well as being (paradoxically) the only project whose paper has never been revised.

I ultimately gathered more than 5,XXX project documents spanning the XXXX projects in the data set, which formed the basis for the corpus I used in the text analysis I describe in Chapter 3 (and below). Translating this large set of documents into a single corpus has proven to be an additional challenge in its own right due to a mismatch between the assumptions around data made by most studies using topic modeling as an analytical approach. These studies have used data that reflects the institutionalized settings within which it was created, whether presidential speeches, regulatory filings or newspaper article databases in which documents are written, stored and presented using unified conventions and a consistent format. In some cases (as with regulatory filings) the documents are so standardized that it is possible to isolate and use a consistent subset of the documents that have only the language required for the analysis. As a result, these studies tend to draw on an established toolkit for cleaning or "rendering" (Hannigan et al. 2019) the documents into a useful set of data, allowing the authors to use and cite those methods without addressing their substance.

Where these traditional corpora are often created in highly institutionalized settings mandating consistent document structures, and stored in file formats with identical conventions, the documents comprising my data set are far more disparate. The proliferation of underlying software used to generate the files, the wide range of countries of origin, formatting conventions, grammatical conventions (or the lack thereof) and inconsistent usage of technical terms together rendered any single rule-based approach to data rendering inadequate.

I address this through a series of steps that are integrally related to the computational methods I develop and use in my analysis. As a first step, I needed to get the documents from .pdf format into the same text file format in order to begin working with the data. This seemed simple enough until I noticed that some texts appeared jumbled when read into R. I discovered that this was due to widely differing formats, some of which involved the templated two-column LaTeX formatting of computer science papers, while others were highly designed and thus had text on some pages in single columns, and in others in three or more. The software library I used to read these documents into R worked mechanically across rows, rendering multicolumn text into an unintended format. After some searching for a solution, I eventually began processing all multicolumn and other structurally challenging documents using a third-party website that rendered them into editable text, which I then manually inspected to identify any errors in the rendering (lines and sometimes full paragraphs could be lost in translating formats, especially when the original file includes images).[2] Dealing with these issues has entailed comparing each of the documents paragraph by paragraph against the source pdf, and adding or editing the Word document as necessary to ensure that it reflects the original ordering of the text.

Doing this manually for somewhere between a quarter and one third of the thousands of white papers in my data set has entailed a level of engagement with the texts themselves has given me a deep familiarity with the structure of both the narratives and the arguments that tend to reappear across projects. It also made me aware of language that is typically included in these documents due to "genre" (Orlikowski and Yates 1994; Yates and Orlikowski 1992) requirements imposed by some combination of the documents' function and their institutional environment. While such language would be helpful for other kinds of analysis, it is extraneous to the classifications I seek to make based on the implementations themselves. The most obvious of these is the legal boilerplate included in the white papers of startup projects seeking to raise funds through a token sale, as well as language describing the terms of the token sale, as well as various supporting information such as the biographies of the founding team. Including technical papers in my data set means that the corpus also includes bibliographies, references to figure names and other characteristics of scientific papers.

I also drew on and modified more traditional methods in working with my data, key among which was identifying words that frequently appeared together and collapsing them into a single word. This was especially important for my corpus given the inconsistent application of hyphens to important terms such as "supply chain," which appeared variously as "supply chain", "supply-chain" and "supplychain." I also applied this approach to other aspects of the corpus that only became evident as I began running topic models. Perhaps the most pernicious (in its subtlety) was the seemingly random presence of word fragments, which I eventually determined was the result of inconsistent spacing around line breaks that only became apparent when I tried to use a single rule around punctuation such as hyphens. I identified these creating a list of all such fragments that frequently appeared together in order to specify the adjacent fragments to join

[2] I primarily used sejda.com for this purpose, and am grateful to Han Zhang for directing me there.

into single tokens (e.g. “decen tralized” but also “decent ralized” both appear as fragments of an important word, as do “sup ply chain” and “supp ly chain”).

I took a similarly customized approach to the process of lemmatizing, or reducing variations on the same word to a single representation. Doing this ensures that the weight of any given word in my data analysis accurately reflects its prominence rather than being dispersed over closely related variants, a move that is particularly important for technical or domain-specific terms that appear relatively less frequently and thus need to have a consistent presence in the modeling. While there are a growing number of established databases or dictionaries for doing this kind of work, I determined after trying several that none was adequate given the nature of my data. I started constructing my own dictionary by identifying the most widely used nouns and verbs, as well as their forms appearing across singular/plural/acronyms for nouns and multiple tenses for verbs by iterating through the data cleaning process, inspecting modeling outputs and using my own knowledge. I applied a similar process for identifying and merging terms using non-American English (a widespread pattern given the global nature of the ecology, and thus of my data set).

I then turned to the more specialized process of collapsing both spelled-out and acronym representations of technical and domain-specific terms into a single representation. The latter was rendered more complex than it sounds given that the same technologies are often described slightly differently. For example, the referent for the acronym HTLC appears in various papers as hashed time-locked contracts, hashed time lock contracts, hashed time lock contracts and hash time-lock contracts, with additional variations generated by the presence/absence and placement of hyphens. Effectively identifying and addressing these terms was only possible after spending considerable time engaged in the algorithmic ethnography I describe above as well as working through the process of transforming two-column technical papers into single-column text.

I added the final form of data cleaning when it became apparent that the method I use for assigning documents to topics (using the topic-document prevalence) was extremely sensitive to the frequency of individual words in ways that aren’t apparent in standard uses of topic modeling.

This was especially the case for words such as "token" and "blockchain" that appeared both as part of proper names and as important terms in their own right. The issue with these terms, I discovered, was that they were often orphaned by the standard approach of removing all terms in a corpus that appear only in some small number of documents, which in my data set effectively removed many proper names but left these terms behind (e.g. "Gnosis token" became "token", an already frequently used term). These were also important in cases where the name of a project (e.g. Diamonds) or the ticker/symbol for its token was also a term that appeared with a different but important meaning in the overall corpus

I addressed these in a similarly customized fashion to the other issues I describe above. While it is possible that these latter instances could have been managed by an off-the-shelf algorithm for identifying proper names, my experience with the heterogeneity of the data and its effects left me leery of attempting to use a single solution given the amount of effort it would have required to verify that the method had worked consistently across thousands of heterogeneous documents. I leave experimentation with part of speech tagging for future work, and instead used the bigram method I describe above to identify multi-word terms based on names (e.g. "Ethereum token"), and continually examined the document-topic prevalence matrix for outliers to identify terms that appeared both formally and in their standard usage (e.g. "Diamonds").

Table A.1: Custom Dictionaries

Library	**Description**	**Elements**
blockdict	Blockchain-specific bigrams and trigrams	638
blockdict	Common phrases in legal disclosure	203
fragprobdict	Frequently split words	845
lemmadict	Lemmatization dictionary	4,660
nameheadfoot	Condensing names with token tickers	1,129
namesdict	Project, organizational names	9,043
postpunctdict	Harmonizing punctuation	26
symboldict	Harmonizing symbols, text	40
ukdict	Harmonizing UK spelling	81

A.5 Developing Taxonomy

The remainder of this Appendix articulates the Methods I describe in Chapter 3 more fully.

A.5.1 Running Topic Models

The standard approach to topic modeling used in the social sciences is to run several models (defined as having different numbers of topics) on the same corpus but to choose only one for interpretation. The decisions involved regarding the number of initial models and the selection criteria for choosing a focal model are mentioned but typically occur offstage. More important, potential insights from the unselected models are also omitted.

My analytical needs lead me to diverge from this single-model approach. Since the goal is to produce a multi-level taxonomy, I need to select multiple models so that each of the three levels of taxonomy has a source, which necessitates running a wide range of models *and* selecting several to use.

I draw insights from two related domains to bound the range of models I run. The NAICS has roughly 20 top-level sectors, which implies a bottom for my range of topic models. By contrast, recent studies of the accuracy of classification using topic models have generally set the number of topics at a multiple of the number of established categories in their data (e.g. Airoldi et al. 2010; Lu, Mei, and Zhai 2011; Xie and Xing 2013). Although I definitionally don't have *ex ante* knowledge of the "true" number of categories, my hand categorizations of the projects in my data set revealed anywhere from 150 to 200 categories, implying that the largest model should be some multiple of that. I thus run a range of topic models using the same corpus but varying the number of topics from 20 to 1300. As an additional input, I also ran several models where I allowed the algorithm to choose the number of topics. The algorithm consistently selected models with between 90-100 topics, with a median of 95. I incorporate three of these models in the analysis that follows.

I generate the individual models within this range using correlated topic models (Blei and

Lafferty 2007), or CTMs. Unlike the Latent Dirichlet Allocation approach used by most analyses using topic models, CTMs do not assume that each topic is independent of the others. Instead, they use a statistical structure that allows for correlations between topics. This makes them both more relevant for my data, given the likelihood of clustering on multiple dimensions, as well as for my analytical question. I use the *stm* package in R without covariates, and have used Spectral initialization to keep my analysis replicable (Roberts et al. 2018). Spectral initialization also facilitates the comparisons across models that inform my labeling.

The final output of this stage of the process was a set of 23 topic models based on the same corpus, but varying in the number of topics used for each model.

A.5.2 Labeling and Coding

I then turn to labeling the topics for each model. Rather than a computational approach, I followed in the tradition of data coding used by qualitative researchers in classifying their data (**grodal_achieving_2020**). Unlike qualitative researchers, however, I am able to capitalize on the variety of weightings and rankings that post-estimation topic modeling algorithms provide. These algorithms take different approaches to the order of presentation of words in a given topic, leading to potentially different interpretations. I capitalize on the *stm* package's flexibility by using all four of the algorithms it offers for probability-weighting words within topics, and use agreement across the four algorithms' output as a strong signal of consistency in a topic's content. In cases where the algorithms don't agree, I relied on interpretation.

As with the initial hand-labeling of projects I used in building my data set, my approach to labeling was an abductive process that I refined over the course of labeling multiple models. Labeling across models also allowed me to identify topics that were uninformative for the analysis. The computer science literature identifies a tradeoff between the benefits of using more topics to get an accurate categorization and the increased likelihood of uninformative topics appearing in larger models (Chuang et al. 2013; Mimno et al. 2011). I follow that literature in terming such topics either *fused* or *junk*, depending on whether they capture two or more than two underlying ideas, respectively. I also identified *genre* topics through this process, meaning those that gathered language from the descriptions of ICO terms, legal boilerplate and other unhelpful language.

A.5.3 Assignment to Classifications

The next question is how to assign documents to labeled topics. Where my approach to labeling or coding reflects qualitative practice, my approach to classification is largely rules-based.

The raw material for my approach is the topic prevalence matrix, which is the model's estimate of the proportion of each document's text generated by each of the topics. There is no standard for what size of prevalence in this range counts as grounds for assigning a document to a topic. I use a simple threshold for assignment of documents to topics that takes account of both the size of a given model as well as the uncertainty or variance of its topic predictions. Developed by Airoldi et al. (2010) for classifying academic articles, this measure sets the assignment threshold at $1/k + sd$, where k is the number of topics and sd is the standard deviation of the posterior of the estimated topic prevalences. Applying this method transforms each model's document-by-topic prevalence matrix of proportions into one in which every entry is either a 1 (indicating that the document-topic prevalence probability exceeded the threshold for inclusion) or a 0 (indicating the converse).

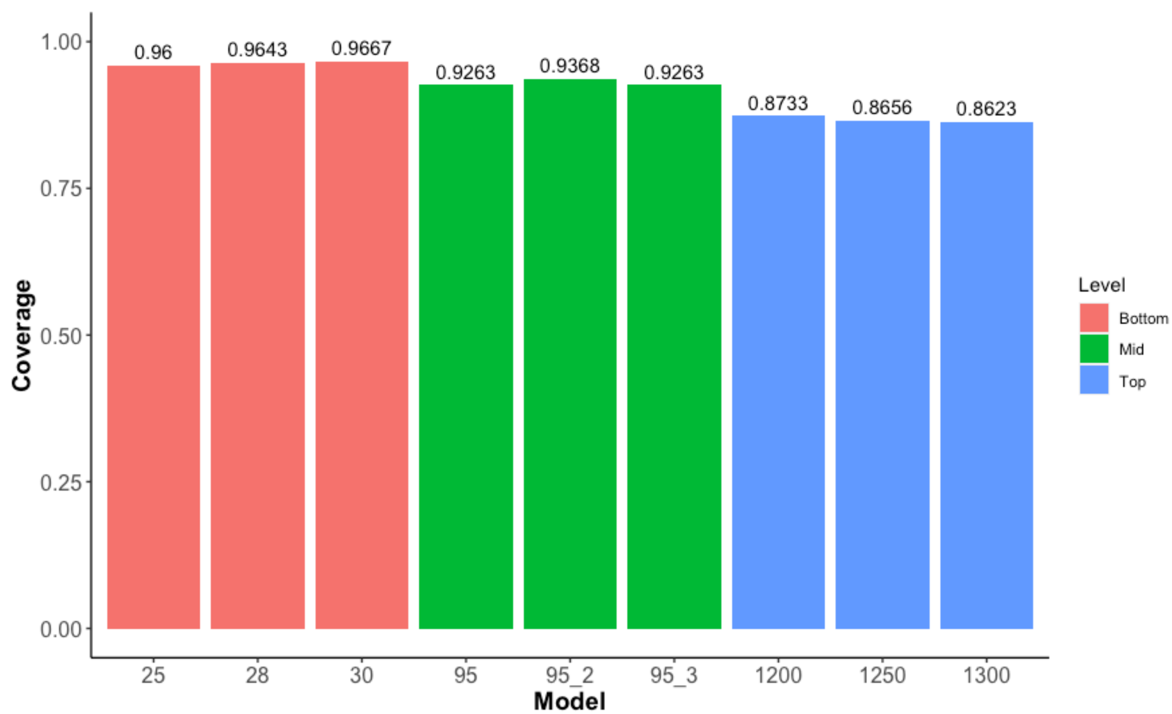
The project-to-topic assignments for each model also give me several means of pruning the output to remove uninformative data. First, the fact that the prevalences are proportions that sum to 1 for each document provides a means of removing the effects of *genre*, *junk* and *fused* topics. I capitalize on this by removing these topics (or, these columns from the prevalence matrix), and then re-renormalizing the prevalence values for the remaining topics across each row to sum to 1. As a control against over-inflation, I eliminate any documents/rows whose aggregate prevalence in the problematic topic vectors exceeded 0.4.

I also use the assignment matrix to address the fact that several hundred projects have multiple documents, which does not align with my treatment of single implementations as my unit of analysis. I address this by merging all the document-topic assignments within a project into a single binary vector.[3]

Finally, the assignments allow me to prune both topics and projects that don't add to the analysis, whether because they are unique or because they provide no information after the stages I describe above. I remove any topics without more than one project assigned, and then remove projects with zero assignments remaining after the prior stages of pruning.

Figure X below

Figure A.2: Topic Retention After Pruning (Coverage)



A.5.4 Selecting Models

Identifying three levels in my approach necessitates selecting models whose sizes (in terms of the number of topics they use) and content vary sufficiently to capture different levels of

[3]More precisely, I take the union of all topic assignments for all documents within a single project, rather than adding them. This preserves the binary nature of the assignments.

information on the same corpus. I take a fundamentally different approach to selection than most studies in that I don't use any single topic as individually representing one of the three levels. Instead, I choose three models at each level. Doing so eliminates the possibility of over-reliance on a single model, and also allows me to refine the labels I've assigned to each topic in light of variations in how those topics appear in similar models.

For maximum contrast, I anchor the top and most general level of the hierarchy in models with among the smallest feasible numbers of topics, and the bottom and most detailed level in models at the other end of the range. This means that the most general level of the hierarchy is based on a combination of models with 25, 28 and 30 topics. The most detailed or bottom level of the hierarchy is in turn based on a combination of models with 1200, 1250 and 1300 topics.

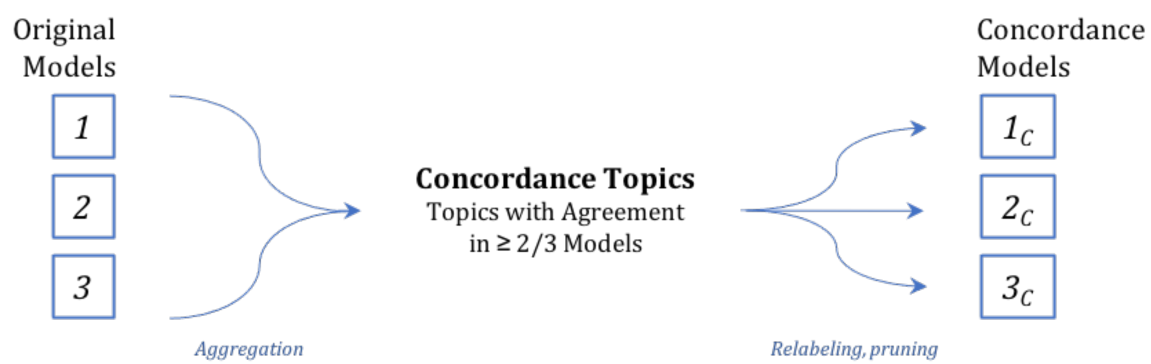
Notably, the three models for which the algorithm chose the number of topics also appeared to mark an inflection point in the output of the models I previously ran (ranging from 20 to 1300 topics). Models with fewer than 95 topics produced identifiably related topics but not in the same order from one model to the next, and with varying emphases. Above it, models settled into a pattern in which topics generally remained where they first appeared until they eventually split into multiple subtopics in some larger model. This pattern leads me to take the models for which the algorithm chose the number of topics as the midpoint of my analysis, meaning that I use them to generate the middle level of the three levels of my category hierarchy.

A.5.5 Establishing Classification Hierarchy

My approach to taxonomy building involves searching for agreement across multiple models (which I term *concordance*) within each of the three levels in order to avoid over-reliance on a single model. I also identify linkages between those levels based on *co-occurrences* in assignments between models across different levels. Finally, I look for *concordance* between the *cooccurrences* to identify those that I interpret as the study's results.

I first search for concordance within levels by identifying the topics that appear in each of the three models in a given level. In order to avoid model-specific results, I require a given topic to appear consistently across at least two of the three models within a given level in order to include it in the subsequent analysis. This involves a careful comparison of content of the topics across models in order to determine both the appropriate topics to compare as well as their shared content. This process produces three sets of *concordance topics* – one for each of the three levels – representing consistent findings across multiple models. I then relabel all the matching topics within each of the three levels to reflect their concordance labels, and remove any topics that did not meet the 2/3 threshold. This process is summarized in Figure C.2.

Figure A.3: Developing Single-Level Concordance Topics



Having established the Top-, Middle- and Detail-Level concordance topics, I then turn to identifying the linkages between levels that define the taxonomy. I start by creating separate cooccurrence matrixes linking each of the three Top-level models to each of the three Middle-Level models, and another six linking individual models in the Middle- and Detail-Level models (see figure x). The rows in these tables are concordance topic-concepts from the larger of the two models, and the columns are concordance topic-concepts from the smaller of the two models. The cells represent the intersections between the row and column in questions, and the numbers in those cells capture the total number of projects assigned to both components of that topic pair.

For example, if three projects were assigned to the "trading" topic in the Middle-Level model and also to the "arbitrage strategies" topic in the Detail-Level model, this would produce a 3 in the cell capturing their intersection in a Middle-/Detail-Level cooccurrence matrix. I would take that entry as evidence of a hierarchical relationship between trading and arbitrage strategies as a component of trading. The more projects that “co-occur” in this way by appearing in a given topic in one model and a second topic in another, the more evidence I find of hierarchical links between them.

As with the concordance topics *within*, I require such co-occurrences to appear in at least two of the three matrixes working *across* levels. The end result is a single set of concordance or consistent linkages defining hierarchical relationships between concordance or consistent concepts at the three levels.

A.5.6 Methodological Limitations

As with any other kind of data, textual data is affected by the potential for biases, especially in cases where the phenomenon they describe has yet to stabilize. The most obvious risk in the case of white papers is the potential for a skewed sample that reflects the convenience of accessing specific kinds of white papers (i.e. those written by startups in the ICO boom of 2017). Although there is no way to be sure to capture every possible implementation, my strategy in gathering papers was to work as broadly as possible in the hope of covering as much of the ecology as possible for this early stage of the technology’s evolution. I did this by attempting to remain current on new papers over the course of my active data gathering, a period that spanned roughly 2017 to early 2020, while also working backward in time to capture any papers that were published prior 2017.

These approaches do little to address the final risk, which the potential for bias given the fact that implementations being created by startups backed by venture capital funding often wait to provide public documentation until they officially launch. I have no formal means of addressing this other than to track high-profile projects as they evolve, and to download their papers when they become available. While far from perfect, this approach at least limits the potential for divergence between my data set and the full population.

There are several other potential threats to the approach I develop here, key among which is the risk inherent in any topic model of over-reliance on a single model, and on the related potential for over-reliance on idiosyncratic interpretations as a result. I have attempted to limit this risk through the incorporation of multiple models, as well as through the process of coding category names by looking across the full spectrum of models (in addition to the various dictionary methods I describe above. There nonetheless remains the final risk of over-reliance on my own interpretation. I look forward to further validating the taxonomy with external audiences as this project continues to develop.

The method I have developed here also points to the potential for future work. Most obviously, the taxonomy in Appendix B could be used as the basis for studying variation in the general structures I develop in Chapter 6 beyond the technological layers I identify there. For example, an ecological analyst could test for equifinality (Hannan and Freeman 1989) in organizational structures across categories. Such heterogeneity could also be identified through a combination of taxonomic and fuzzy set analysis.

The scope of the data set developed for this project also opens the door to a rich spectrum of potential work. For example, it could be possible to develop a text-based measure of complementarities in digital organization.

References

Abbott, Andrew. 2016. *Processual Sociology*. Chicago: The University of Chicago Press.

Ackerman, Bruce. 1998. *We the People, Volume 2: Transformations*. Harvard University Press.

Adler, Paul, and Charles Heckscher. 2007. “Towards collaborative community”. In *The collaborative enterprise: Managing speed and complexity in knowledge-based businesses*, ed. by Charles Heckscher and Paul Adler. New Haven: Yale University Press.

Adner, Ron. 2017. “Ecosystem as Structure: An Actionable Construct for Strategy”. *Journal of Management* 43 (1): 39–58.

Adner, Ron, and Rahul Kapoor. 2010. “Value Creation in Innovation Ecosystems: How the Structure of Technological Interdependence Affects Firm Performance in New Technology Generations”. *Strategic Management Journal* 31 (3): 306–333.

Agrawal, Ajay, Christian Catalini, and Avi Goldfarb. 2014. “Some Simple Economics of Crowdfunding”. *Innovation Policy and the Economy* 14:63–97.

Agrawal, Ajay K, Christian Catalini, and Avi Goldfarb. 2011. *The Geography of Crowdfunding*. Working Paper 16820. National Bureau of Economic Research.

Ahrne, Göran, and Nils Brunsson. 2011. “Organization outside organizations: the significance of partial organization”. *Organization* 18 (1): 83–104.

Airoldi, Edoardo M., et al. 2010. “Reconceptualizing the classification of PNAS articles”. *Proceedings of the National Academy of Sciences* 107 (49): 20899–20904.

Ansell, Christopher K. 2011. *Pragmatist democracy: evolutionary learning as public philosophy*. New York: Oxford University Press.

Aoki, Masahiko. 1994. “The Contingent Governance of Teams: Analysis of Institutional Complementarity”. *International Economic Review* 35 (3): 657–676.

Arazy, Ofer, Hila Lifshitz□Assaf, and Adam Balila. 2019. “Neither a Bazaar nor a Cathedral: The Interplay Between Structure and Agency in Wikipedia’s Role System”. *Journal of the Association for Information Science and Technology* 70 (1): 3–15.

Arluck, Jacob. 2018. “Amending Tezos: Traversing the Amendment Process”. Medium. Library Catalog: medium.com. Visited on 03/04/2020. https://medium.com/tezos/amending-tezos-b77949d97e1e.

Arrow, Kenneth J. 1950. “A Difficulty in the Concept of Social Welfare”. *Journal of Political Economy* 58 (4): 328–346.

Auer, Raphael, and Rainer Böhme. 2020. *The Technology of Retail Central Bank Digital Currency*. Bank for International Settlements.

Auer, Raphael, Giulio Cornelli, and Jon Frost. 2020. *Rise of the Central Bank Digital Currencies: Drivers, Approaches and Technologies*. CESIFO Working Paper 8655. Munich: Center for Economic Studies and Ifo Institute (CESifo).

Back, Adam. 1997. “Hash Cash Postage Implementation”. Cypherpunk listserv. http://www.hashcash.org/papers/announce.txt.

Baden-Fuller, Charles, and Mary S. Morgan. 2010. “Business Models as Models”. *Long Range Planning*, Business Models, 43 (2): 156–171.

Baldwin, Carliss Y., and Kim B. Clark. 2006. “The Architecture of Participation: Does Code Architecture Mitigate Free Riding in the Open Source Development Model?” *Management Science* 52 (7): 1116–1127.

Bandelj, Nina, Frederick F. Wherry, and Viviana A. Zelizer. 2017. “Introduction. Advancing Money Talks”. In *Money Talks: Explaining How Money Really Works*, ed. by Nina Bandelj, Frederick F. Wherry, and Viviana A. Zelizer, 1–22. Princeton University Press. Visited on 09/05/2021. doi:10.1515/9781400885268-003.

Barley, Stephen R. 1990. “The alignment of technology and structure through roles and networks”. *Administrative Science Quarterly* 35 (1): 61–103.

Barrett, Michael, Eivor Oborn, and Wanda Orlikowski. 2016. “Creating Value in Online Communities: The Sociomaterial Configuring of Strategy, Platform, and Stakeholder Engagement”. *Information Systems Research* 27 (4): 704–723.

Bechky, Beth A. 2006. “Gaffers, Gofers, and Grips: Role-Based Coordination in Temporary Organizations”. *Organization Science* 17 (1): 3–21.

Beck, Roman, Christoph Müller-Bloch, and John Leslie King. 2018. “Governance in the Blockchain Economy: A Framework and Research Agenda”. *Journal of the Association for Information Systems* 19 (10): 1020.

Beckert, Jens. 1999. “Agency, Entrepreneurs, and Institutional Change. The Role of Strategic Choice and Institutionalized Practices in Organizations”. *Organization Studies* 20 (5): 777–799.

– . 2016. *Imagined Futures*. Cambridge, MA: Harvard University Press.

Bekar, Clifford, Kenneth Carlaw, and Richard Lipsey. 2018. “General purpose technologies in theory, application and controversy: a review”. *Journal of Evolutionary Economics* 28 (5): 1005–1033.

Benkler, Yochai. 2002. “Coase’s Penguin, or, Linux and ”The Nature of the Firm””. *The Yale Law Journal* 112 (3): 369–446.

– . 2013. “Commons and Growth: The Essential Role of Open Commons in Market Economies”. Ed. by Brett M. Frischmann. *The University of Chicago Law Review* 80 (3): 1499–1555.

– . 2016. “Degrees of Freedom, Dimensions of Power”. *Daedalus* 145, no. 1 (): 18–32.

– . 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven: Yale University Press.

Benoit, Kenneth, et al. 2018. “quanteda: Quantitative Analysis of Textual Data”. https://CRAN.R-project.org/package=quanteda.

Berg, Chris, Sinclair Davidson, and Jason Potts. 2019. *Understanding the Blockchain Economy: An Introduction to Institutional Cryptoeconomics*. New horizons in institutional and evolutionary economics. Northampton: Edward Elgar Publishing.

Birch, Kean. 2017. “Rethinking Value in the Bio-economy: Finance, Assetization, and the Management of Value”. *Science, Technology, & Human Values* 42 (3): 460–490.

Birch, Kean, and Fabian Muniesa, eds. 2020. *Assetization: Turning Things into Assets in Technoscientific Capitalism*. Cambridge, MA: MIT Press.

Blair, Margaret M., and Lynn A. Stout. 1999. “A Team Production Theory of Corporate Law”. *Virginia Law Review* 85 (2): 247–328.

– . 2005. “Specific Investment: Explaining Anomalies in Corporate Law”. *Journal of Corporation Law* 31:719. https://heinonline.org/HOL/Page?handle=hein.journals/jcorl31&id=731&div=&collection=.

Blau, Peter M. 1963. *The dynamics of bureaucracy; a study of interpersonal relations in two Government agencies*. Rev. [2d] ed. Chicago: University of Chicago Press.

Blei, David M., and John D. Lafferty. 2007. “A correlated topic model of science”. *The Annals of Applied Statistics* 1 (1): 17–35.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. “Latent Dirichlet Allocation”. *Journal of Machine Learning Research* 3 (Jan): 993–1022.

Borrás, Susana. 2012. “Three Tensions in the Governance of Science and Technology”. In *The Oxford Handbook of Governance*, ed. by David Levi-Faur, 429–440. Oxford: Oxford University Press.

Boudreau, Kevin J., and Karim R. Lakhani. 2015. "'Open' disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology". *Research Policy* 44 (1): 4–19.

Bowker, Geoffrey C. 1994. *Science on the run: information management and industrial geophysics at Schlumberger, 1920-1940*. Inside technology. Cambridge, MA: MIT Press.

Brandtner, Christof. 2017. "Putting the World in Orders: Plurality in Organizational Evaluation". *Sociological Theory* 35 (3): 200–227.

Bresnahan, Timothy F., and M. Trajtenberg. 1995. "General Purpose Technologies: Engines of Growth?" *Journal of Econometrics* 65 (1): 83–108.

Brown, John Seely, and Paul Duguid. 1991. "Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation". *Organization Science* 2 (1): 40–57.

Brunton, Finn. 2019. *Digital Cash: The Unknown History of the Anarchists, Utopians, and Technologists who Built Cryptocurrency*. Princeton, N.J.: Princeton University Press.

Burton, Richard M., and Børge Obel. 2013. "The science of organizational design: fit between structure and coordination". *Journal of Organization Design* 7 (1): 5.

Buterin, Vitalik. 2013. *A next generation smart contract & decentralized application platform*. Ethereum White Paper. Ethereum. ethereum.org.

– . 2017. "Notes on Blockchain Governance". Vitalik Buterin's website. Visited on 03/09/2019. https://vitalik.ca/general/2017/12/17/voting.html.

Caliskan, Koray. 2020. "Data Money: The Socio-Technical Infrastructure of Cryptocurrency Blockchains". *Economy and Society* 49 (4): 540–561.

Callon, Michel. 1984. "Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay". *The Sociological Review* 32 (): 196–233.

Caramani, Daniele. 2017. “Will vs. Reason: The Populist and Technocratic Forms of Political Representation and Their Critique to Party Government”. *The American Political Science Review; Washington* 111 (1): 54–67.

Carlaw, Kenneth, and Richard G. Lipsey. 2002. “Externalities, Technological Complementarities and Sustainable Economic Growth”. *Research Policy* 31 (8): 1305–1315.

Carruthers, Bruce G., and Jeong-Chul Kim. 2011. “The sociology of finance”. *Annual Review of Sociology* 37 (1): 239–259.

Carter, Nic. 2018. “Blockchain Is a Semantic Wasteland”. Medium. https://medium.com/s/story/blockchain-is-a-semantic-wasteland-9450b6e5012.

Castells, Manuel. 2011. “A Network Theory of Power”. *International Journal of Communication* 5 (773): 15.

Castro, Miguel, and Barbara Liskov. 1999. “Practical Byzantine Fault Tolerance”. In *Proceedings of the Third Symposium on Operating Systems Design and Implementation*, 173–186. Visited on 05/22/2019.

Catalini, Christian, and Joshua S. Gans. 2016. *Some Simple Economics of the Blockchain*. Working Paper 22952. National Bureau of Economic Research.

Cedefop. 2020. *Developing and Matching Skills in the Online Platform Economy: Findings on New Forms of Digital Work and Learning from Cedefop’s CrowdLearn Study* Cedefop reference series: no. 116. Luxembourg: Publications Office: European Union.

Chaum, David, Christian Grothoff, and Thomas Moser. 2021. *How to Issue a Central Bank Digital Currency* Issue 2021-03. Zurich, Switzerland: Swiss National Bank. Visited on 04/05/2021. https://www.snb.ch/en/mmr/papers/id/working_paper_2021_03.

Christin, Angèle. 2020. “The ethnographer and the algorithm: beyond the black box”. *Theory and Society* 49 (5): 897–918.

Chuang, Jason, et al. 2013. “Topic Model Diagnostics: Assessing Domain Relevance via Topical Alignment”. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, III–612–III–620. ICML’13. Atlanta, GA, USA: JMLR.org.

Citron, Danielle Keats. 2007. “Technological Due Process”. *Washington University Law Review* 85:1249.

Clack, Christopher D., Vikram A. Bakshi, and Lee Braine. 2016. “Smart Contract Templates: foundations, design landscape and research directions”. *arXiv:1608.00771 [cs]* ().

Colfer, Lyra J., and Carliss Y. Baldwin. 2016. “The mirroring hypothesis: theory, evidence, and exceptions”. *Industrial and Corporate Change* 25 (5): 709–738.

Colyvas, Jeannette A., and Stefan Jonsson. 2011. “Ubiquity and Legitimacy: Disentangling Diffusion and Institutionalization”. *Sociological Theory* 29 (1): 27–53.

Cong, Lin William, Ye Li, and Neng Wang. 2019a. *Tokenomics and Platform Finance*. SSRN Scholarly Paper ID 3472481. Rochester, NY: Social Science Research Network. Visited on 10/30/2019. https://papers.ssrn.com/abstract=3478193.

– . 2019b. *Tokenomics: Dynamic Adoption and Valuation*. SSRN Scholarly Paper ID 3153860. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=3153860.

Conway, Melvin E. 1968. “How do committees invent?” *Datamation* 14 (5): 28–31.

Coulouris, George, et al. 2012. *Distributed Systems: Concepts and Design*. 5th. Boston: Addison-Wesley.

Dai, Wei. 1998. “B-Money, An Autonomous, Distributed Electronic Cash System”.

Daniels, Jessie, et al., eds. 2017. *Digital Sociologies*. Reprint edition. Bristol, UK Chicago, IL: Policy Press.

De Filippi, Primavera, and Aaron Wright. 2018. *Blockchain and the Law: The Rule of Code*. Cambridge: Harvard University Press.

DeLanda, Manuel. 2006. *A new philosophy of society: assemblage theory and social complexity*. London: Continuum.

DeNardis, Laura, and Andrea M. Hackl. 2015. “Internet Governance by Social Media Platforms”. *Telecommunications Policy* 39 (9): 761–770.

Desan, Christine. 2014. *Making Money: Coin, Currency and the Coming of Capitalism*. Oxford: Oxford University Press.

Dhanaraj, Charles, and Arvind Parkhe. 2006. “Orchestrating Innovation Networks”. *Academy of Management Review* 31 (3): 659–669.

Diekmann, Andreas, et al. 2014. “Reputation Formation and the Evolution of Cooperation in Anonymous Online Markets”. *American Sociological Review* 79 (1): 65–85. ISSN: 0003-1224, 1939-8271, visited on 01/03/2016.

DiMaggio, Paul J., and Walter W. Powell. 1983. “The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields”. *American Sociological Review* 48 (2): 147–160.

Dobusch, Leonhard, and Dennis Schoeneborn. 2015. “Fluidity, Identity, and Organizationality: The Communicative Constitution of Anonymous”. *Journal of Management Studies* 52 (8): 1005–1035.

Dodd, Nigel. 2014. *The social life of money*. Princeton, New Jersey: Princeton University Press.

– . 2017. “Utopian moneys: complementary currencies, bitcoin, and the social life of money”. In *Money Talks*, ed. by Nina Bandelj, Frederick F. Wherry, and Viviana A. Zelizer, 230–247. Princeton: Princeton University Press.

Doganova, Liliana, and Marie Eyquem-Renault. 2009. “What do business models do?: Innovation devices in technology entrepreneurship”. *Research Policy* 38 (10): 1559–1570.

DuPont, Quinn. 2018. “Experiments in Algorithmic Governance: A History and Ethnography of ”The DAO,” a Failed Decentralized Autonomous Organization”. In *Bitcoin and Beyond Cryptocurrencies, Blockchains, and Global Governance*, ed. by Malcolm Campbell-Verduyn, 157–177. New York: Routledge.

Durand, Rodolphe, and Mukti Khaire. 2017. “Where do market categories come from and how? Distinguishing category creation from category emergence”. *Journal of Management* 43 (1): 87–110.

Durand, Rodolphe, and Lionel Paolella. 2013. “Category Stretching: Reorienting Research on Categories in Strategy, Entrepreneurship, and Organization Theory”. *Journal of Management Studies* 50 (6): 1100–1123.

Durkheim, Emile. 1997. *The Division of Labor in Society*. Trans. by W.D. Halls. Simon / Schuster.

Dyson, Esther, et al. 1996. “Cyberspace and the American dream: a Magna Carta for the knowledge age”. *The Information Society* 12 (3): 295–308.

Eaton, Ben, et al. 2015. “Distributed Tuning of Boundary Resources: The Case of Apple’s iOS Service System”. *Management Information Systems Quarterly* 39 (1): 217–243.

Economist. 2015. “The trust machine”. The Economist. Visited on 11/15/2016. http://www.economist.com/news/leaders/21677198-technology-behind-bitcoin-could-transform-how-economy-works-trust-machine.

– . 2017. “The world’s most valuable resource is no longer oil, but data”. *The Economist* (). Visited on 01/18/2019. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

Edgerton, David. 2007. *The shock of the old: technology and global history since 1900*. New York: Oxford University Press.

Eghbal, Nadia. 2016. *Roads and Bridges: The Unseen Labor Behind our Digital Infrastructure*. Ford Foundation. https://www.fordfoundation.org/about/library/reports-and-studies/roads-and-bridges-the-unseen-labor-behind-our-digital-infrastructure.

– . 2020. *Working in Public: The Making and Maintenance of Open Source Software*. Stripe Press.

Ehrsam, Fred. 2017. “Blockchain Governance: Programming Our Future”. Fred Ehrsam. Visited on 03/09/2019. https://medium.com/@FEhrsam/blockchain-governance-programming-our-future-c3bfe30f2d74.

Eisenhardt, Kathleen M., and Melissa E. Graebner. 2007. “Theory Building From Cases: Opportunities And Challenges”. *Academy of Management Journal* 50 (1): 25–32.

Elster, Jon, ed. 1998. *Deliberative Democracy*. Cambridge studies in the theory of democracy. Cambridge: Cambridge University Press.

Emirbayer, Mustafa, and Ann Mische. 1998. “What Is Agency?” *American Journal of Sociology* 103 (4): 962–1023.

Evan, William M., and Paul Olk. 1990. “R&D Consortia: A New U.S. Organizational Form”. *Sloan Management Review* 31 (3): 37–46.

Feldman, Maryann P., and Ji Woong Yoon. 2012. “An Empirical Test for General Purpose Technology: An Examination of the Cohen–Boyer rDNA Technology”. *Industrial and Corporate Change* 21 (2): 249–275.

Felin, Teppo, and Todd R. Zenger. 2017. “The theory-based view: Economic actors as theorists”. *Strategy Science* 2 (4): 258–271.

Fishkin, James S. 2011. “The Trilemma of Democratic Reform”. In *When the People Speak: Deliberative Democracy and Public Consultation*, 32–64. Oxford; New York: Oxford University Press.

Ford, Bryan. 2002. *Delegative Democracy*.

Fourcade, Marion. 2017. “The fly and the cookie: alignment and unhingement in 21st-century capitalism”. *Socio-Economic Review* 15 (3): 661–678.

Fourcade, Marion, and Kieran Healy. 2017. “Seeing Like a Market”. *Socio-Economic Review* 15 (1): 9–29.

Fox, Mark S. 1981. “An Organizational View of Distributed Systems”. *IEEE Transactions on Communication Systems, Man and Cybernetics* SMC-11 (1): 70–80.

Freeman, Jo. 1972. “The Tyranny of Structurelessness”. *Berkeley Journal of Sociology* 17:151–164.

Frolov, Daniil. 2021. “Blockchain and institutional complexity: an extended institutional approach”. *Journal of Institutional Economics* 17 (1): 21–36. Visited on 04/05/2021.

Fu, Yondon. 2020. “Livepeer Governance Roadmap Proposal”. Livepeer Blog. Visited on 03/23/2020. https://medium.com/livepeer-blog/livepeer-governance-roadmap-proposal-69a6e9e33f80.

Gagolewski, Marek, et al. 2018. “stringi: Character String Processing Facilities”. https://CRAN.R-project.org/package=stringi.

Galloway, Alexander R. 2004. *Protocol: How Control Exists After Decentralization*. Cambrige, MA: MIT Press.

Gawer, Annabelle. 2014. “Bridging Differing Perspectives on Technological Platforms: Toward an Integrative Framework”. *Research Policy* 43 (7): 1239–1249.

Geiger, R Stuart. 2017. “Beyond opening up the black box: Investigating the role of algorithmic systems in Wikipedian organizational culture”. *Big Data & Society* 4 (2): 1–14.

Gerstner, Louis V. 2002. *Who Says Elephants Can’t Dance?: Inside IBM’s Historic Turnaround*. New York: HarperCollins.

Ghazawneh, Ahmad, and Ola Henfridsson. 2013. “Balancing platform control and external contribution in third-party development: the boundary resources model”. *Information Systems Journal* 23 (2): 173–192.

Ghaziani, Amin, and Marc J. Ventresca. 2005. “Keywords and Cultural Change: Frame Analysis of Business Model Public Talk, 1975-2000”. *Sociological Forum* 20 (4): 523–559.

Giddens, Anthony. 1990. *The Consequences of Modernity*. Google-Books-ID: oU99QgAACAAJ. Stanford University Press.

Gilbert, S., and N. Lynch. 2012. “Perspectives on the CAP Theorem”. *Computer* 45, no. 2 (): 30–36.

Gilbert, Seth, and Nancy Lynch. 2002. “Brewer’s conjecture and the feasibility of consistent, available, partition-tolerant web services”. *SIGACT News* 33 (2): 51–59.

Giudici, Alessandro, Patrick Reinmoeller, and Davide Ravasi. 2017. “Open-System Orchestration as a Relational Source of Sensing Capabilities: Evidence from a Venture Association”. *Academy of Management Journal* 61 (4): 1369–1402.

Glaser, Barney G., and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Observations. Chicago: Aldine Pub. Co.

Glaser, Vern L., Mariam Krikorian Atkinson, and Peer C. Fiss. 2019. “Goal-Based Categorization: Dynamic Classification in the Display Advertising Industry”. *Organization Studies*: 1–23.

Glynn, Mary Ann, Pamela s. Barr, and M. Tina Dacin. 2000. “Pluralism and the Problem of Variety”. *Academy of Management Review* 25 (4): 726–734.

Goetzmann, William N. 2016. *Money Changes Everything: How Finance Made Civilization Possible*. Princeton University Press.

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. Doubleday anchor books, A174. Garden City, N.Y.: Doubleday.

Goldfarb, Avi, Bleda Taska, and Florenta Teodoridis. 2020. “Could Machine Learning Be a General Purpose Technology? Evidence from Online Job Postings”. University of Toronto Working Paper.

Goldin, Mike. 2017. “Token-Curated Registries 1.0”. Medium. Visited on 04/24/2018. https://medium.com/@ilovebagels/token-curated-registries-1-0-61a232f8dac7.

Gonella, Theo. 2020. “0x Governance in 2020 and Beyond”. 0x Blog. https://web.archive.org/web/20200311194040/https://blog.0xproject.com/0x-governance-in-2020-and-beyond-9e7703bc1c54?gi=cb874f86be58.

Goodman, L.M. 2014. *Tezos - A Self-Amending Crypto-Ledger*. White Paper. https://tezos.com/static/white_paper-2dc8c02267a8fb86bd67a108199441bf.pdf.

Grabher, Gernot. 2004. “Temporary Architectures of Learning: Knowledge Governance in Project Ecologies”. *Organization Studies* 25 (9): 1491–1514.

– . 2002. “The Project Ecology of Advertising: Tasks, Talents and Teams”. *Regional Studies* 36 (3): 245–262.

Grandori, Anna, and Santi Furnari. 2008. “A Chemistry of Organization: Combinatory Analysis and Design”. *Organization Studies* 29 (3): 459–485.

– . 2013. “Configurational Analysis and Organization Design: Towards a Theory of Structural Heterogeneity”. In *Configurational Theory and Methods in Organizational Research*, ed. by Peer C. Fiss, Bart Cambré, and Axel Marx, 38:77–105. Research in the Sociology of Organizations. Emerald Group Publishing Limited.

Granqvist, Nina, Stine Grodal, and Jennifer L. Woolley. 2012. “Hedging Your Bets: Explaining Executives’ Market Labeling Strategies in Nanotechnology”. *Organization Science* 24 (2): 395–413.

Greif, Avner. 1989. “Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders”. *The Journal of Economic History* 49 (4): 857–882. Visited on 12/15/2017.

Grigg, Ian. 2000. “Financial Cryptography in 7 Layers”. In *Financial Cryptography Fourth International Conference*. Anguilla, British West Indies. http://iang.org/papers/fc7.html.

– . 2004. “The Ricardian Contract”, 25–31. IEEE. http://iang.org/papers/ricardian_contract.html.

– . 2005. *Triple Entry Accounting* Rev. 1.7. Visited on 02/02/2020. https://iang.org/papers/triple_entry.html.

Grimmelmann, James. 2005. “Regulation by Software”. *The Yale Law Journal* 114 (7): 1719–1758.

Grodal, Stine, Aleksios Gotsopoulos, and Fernando F. Suarez. 2015. “The Coevolution of Technologies and Categories During Industry Emergence”. *Academy of Management Review* 40 (3): 423–445.

Gulati, Ranjay, Phanish Puranam, and Michael Tushman. 2012. “Meta-organization design: Rethinking design in interorganizational and community contexts”. *Strategic Management Journal* 33 (6): 571–586.

Habermas, Jürgen. 1989. *The Structural Transformation of the Public Sphere: An Inquiry Into a Category of Bourgeois Society*. Trans. by Thomas Burger. Cambridge MA: MIT Press.

Halaburda, Hanna. 2016. “Digital Currencies: Beyond Bitcoin”. *Digiworld Economic Journal*, no. 103: 77–92.

Halaburda, Hanna, and Miklos Sarvary. 2016. *Beyond Bitcoin: The Economics of Digital Currencies*. Palgrave Macmillan.

Hall, Bronwyn H, and Manuel Trajtenberg. 2004. *Uncovering GPTs with Patent Data*. Working Paper 10901. National Bureau of Economic Research. Visited on 07/20/2020.

Hall, Peter A., and David W. Soskice. 2001. “An Introduction to Varieties of Capitalism”. In *Varieties of capitalism: the institutional foundations of comparative advantage*, ed. by Peter A. Hall and David W. Soskice. Oxford; New York: Oxford University Press.

Hannan, Michael T., and John Freeman. 1989. *Organizational Ecology*. Cambridge, Mass: Harvard University Press.

– . 1977. “The population ecology of organizations”. *American Journal of Sociology* 82 (5): 929–964.

Hannan, Michael T., et al. 2019. *Concepts and Categories: Foundations for Sociological and Cultural Analysis*. First. New York: Columbia University Press.

Hannigan, Timothy R., et al. 2019. “Topic Modeling in Management Research: Rendering New Theory from Textual Data”. *Academy of Management Annals* 13 (2): 586–632.

Hanson, Robin. 2013. “Shall We Vote on Values, But Bet on Beliefs?” *Journal of Political Philosophy* 21 (2): 151–178.

Haraway, Donna J. 2016. *Staying with the Trouble: Making Kin in the Chthulucene*. Duke University Press.

Hardin, Garrett. 1985. *Filters against Folly: How to Survive Despite Economists, Ecologists and the Merely Eloquent*. New York, N.Y: Viking Adult.

Hess, Charlotte, and Elinor Ostrom. 2003. “Ideas, Artifacts, and Facilities: Information as a Common-Pool Resource”. *Law and Contemporary Problems* 66 (1): 111–145.

– . 2007. *Understanding Knowledge as a Commons: From Theory to Practice*. Cambridge, MA: MIT Press.

Hsieh, Ying-Ying, et al. 2018. “Bitcoin and the rise of decentralized autonomous organizations”. *Journal of Organization Design* 7 (1): 14.

Hui, Julie S., Michael D. Greenberg, and Elizabeth M. Gerber. 2014. “Understanding the role of community in crowdfunding work”. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 62–74. CSCW ’14. Baltimore, Maryland, USA: Association for Computing Machinery. Visited on 07/24/2020.

Iansiti, Marco, and Karim R. Lakhani. 2017. “The truth about blockchain”. *Harvard Business Review* 95 (1): 118–127.

Ingram, Paul, and Magnus Thor Torfason. 2010. “Organizing the in-Between: The Population Dynamics of Network-Weaving Organizations in the Global Interstate Network”. *Administrative Science Quarterly* 55 (4): 577–605.

Jackson, Matthew O. 2013. *Matching, Auctions, and Market Design*. SSRN Scholarly Paper ID 2263502. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2263502.

– . 2014. *Mechanism Theory*. SSRN Scholarly Paper ID 2542983. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2542983.

Jacob, François. 1982. *The Possible and the Actual*. Seattle, WA: University of Washington Press.

Jacobides, Michael G., Carmelo Cennamo, and Annabelle Gawer. 2018. "Towards a Theory of Ecosystems". *Strategic Management Journal* 39 (8): 2255–2276.

Johnson, Victoria. 2007. "What Is organizational imprinting? Cultural entrepreneurship in the founding of the Paris Opera". *American Journal of Sociology* 113 (1): 97–127.

Katz, Jerome, and William B. Gartner. 1988. "Properties of Emerging Organizations". *Academy of Management Review* 13, no. 3 (): 429–441.

Katz, Jonathan, and Yehuda Lindell. 2015. *Introduction to Modern Cryptography*. Vol. Second edition. Chapman & Hall/CRC cryptography and network security. Boca Raton, Florida: CRC Press, Taylor & Francis Group.

Kellogg, Katherine, Melissa Valentine, and Angele Christin. 2019. "Algorithms at Work: The New Contested Terrain of Control". *Academy of Management Annals*.

Kennedy, Mark Thomas. 2008. "Getting Counted: Markets, Media, and Reality". *American Sociological Review* 73 (2): 270–295.

Kennedy, Mark Thomas, Jade (Yu-Chieh) Lo, and Michael Lounsbury. 2010. "Category currency: the changing value of conformity as a function of ongoing meaning construction". In *Categories in Markets: Origins and Evolution*, ed. by Greta Hsu, Giacomo Negro, and Ozgecan Kocak, 31:369–397. Research in the Sociology of Organizations. Bingley, UK: Emerald Group Publishing Limited.

Khaire, Mukti, and R. Daniel Wadhwani. 2010. “Changing landscapes: the construction of meaning and value in a new market category—modern Indian art”. *Academy of Management Journal* 53 (6): 1281–1304.

Klein, Peter G., et al. 2019. “Organizational Governance Adaptation: Who Is In, Who Is Out, and Who Gets What”. *Academy of Management Review* 44 (1): 6–27.

Klijn, Erik-Hans, and Chris Skelcher. 2007. “Democracy and Governance Networks: Compatible or Not?” *Public Administration* 85 (3): 587–608.

Knorr Cetina, Karin. 1981. “Introduction”. In *Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-Sociologies*, ed. by Karin Knorr Cetina and Aaron Cicourel, 1–30. London: Routledge.

– . 1997. “Sociality with Objects: Social Relations in Postsocial Knowledge”. *Theory, Culture & Society* 14 (4): 1–30.

Kogut, Bruce. 2000. “The Network as Knowledge: Generative Rules and the Emergence of Structure”. *Strategic Management Journal* 21 (3): 405–425. http://www.jstor.org/stable/3094194.

Koning, J.P. 2020. “Programmable money isn’t new, we’ve had it for ages”. Moneyness. http://jpkoning.blogspot.com/2020/11/programmable-money-isnt-new-weve-had-it.html.

Kuwabara, Ko. 2015. “Do Reputation Systems Undermine Trust? Divergent Effects of Enforcement Type on Generalized Trust and Trustworthiness”. *American Journal of Sociology* 120 (5): 1390–1428.

Lakhani, Karim R, Hila Lifshitz-Assaf, and Michael L. Tushman. 2013. “Open Innovation and Organizational Boundaries: Task Decomposition, Knowledge Distribution and the Locus of Innovation”. In *Handbook of Economic Organization: Integrating Economic and Organization Theory*, ed. by Anna Grandori. Cheltenham, UK ; Northampton, MA: Edward Elgar.

Lalley, Steven P., and E. Glen Weyl. 2018. “Quadratic Voting: How Mechanism Design Can Radicalize Democracy”. *AEA Papers and Proceedings* 108:33–37.

Lamoreaux, Naomi R. 2006. “Did Insecure Property Rights Slow Economic Development? Some Lessons from Economic History”. *Journal of Policy History* 18 (1): 146–164.

Lamport, Leslie. 1998. “The Part-time Parliament”. *ACM Trans. Comput. Syst.* 16 (2): 133–169.

– . 1978. “Time, Clocks, and the Ordering of Events in a Distributed System”. *Commun. ACM* 21 (7): 558–565.

Lamport, Leslie, Robert Shostak, and Marshall Pease. 1982. “The Byzantine Generals Problem”. *ACM Trans. Program. Lang. Syst.* 4 (3): 382–401.

Lanier, Jaron. 2018. *Ten Arguments for Deleting Your Social Media Accounts Right Now*. New York: Henry Holt & Co.

Latour, Bruno. 2005. *Reassembling the social: an introduction to actor-network-theory*. Clarendon lectures in management studies. Oxford; New York: Oxford University Press.

– . 1987. *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, Mass: Harvard University Press.

– . 1990. “Technology is society made durable”. *The Sociological Review* 38 (S1): 103–131.

Lave, Jean, and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.

Lawrence, Paul R., and Jay W. Lorsch. 1967. “Differentiation and integration in complex organizations”. *Administrative Science Quarterly* 12 (1): 1–47.

Lee, Alexander. 2021. “What is programmable money?” FEDS Notes. Visited on 08/12/2021. https://www.federalreserve.gov/econres/notes/feds-notes/what-is-programmable-money-20210623.htm.

Leonardi, Paul M. 2009. “Crossing the Implementation Line: The Mutual Constitution of Technology and Organizing Across Development and Use Activities”. *Communication Theory* 19 (3): 278–310.

– . 2011. “When Flexible Routines Meet Flexible Technologies: Affordance, Constraint, and the Imbrication of Human and Material Agencies”. *MIS Quarterly* 35 (1): 147–167.

Lerner, Josh, and Jean Tirole. 2002. “Some Simple Economics of Open Source”. *The Journal of Industrial Economics* 50 (2): 197–234.

– . 2005. “The Economics of Technology Sharing: Open Source and Beyond”. *The Journal of Economic Perspectives* 19 (2): 99–120.

– . 2001. “The open source movement: Key research questions”. *European Economic Review* 45 (4): 819–826.

Lessig, Lawrence. 2006. *Code: And Other Laws of Cyberspace, Version 2.0*. 2nd Revised ed. New York: Basic Books.

Lester, Richard K., and Michael J. Piore. 2004. *Innovation - the Missing Dimension*. London: Harvard University Press.

Levi-Faur, David. 2012. “From “Big Government” to “Big Governance”?” In *The Oxford Handbook of Governance*, ed. by David Levi-Faur, 3–18. Oxford: Oxford University Press.

Levy, Karen E. C. 2017. “Book-smart, not street-smart: blockchain-based smart contracts and the social workings of law”. *Engaging Science, Technology, and Society* 3 (0): 1–15.

Lipsey, Richard G., Cliff Bekar, and Kenneth Carlaw. 1998. “What Requires Explanation?” In *General Purpose Technologies and Economic Growth*, ed. by Elhanan Helpman, 15–54. Cambridge, Mass: MIT Press.

List, Christian. 2011. “The Logical Space of Democracy”. *Philosophy & Public Affairs* 39 (3): 262–297.

Lounsbury, Michael, and Mary Ann Glynn. 2001. “Cultural entrepreneurship: stories, legitimacy, and the acquisition of resources”. *Strategic Management Journal* 22 (6): 545–564.

Lu, Yue, Qiaozhu Mei, and ChengXiang Zhai. 2011. “Investigating task performance of probabilistic topic models: an empirical study of PLSA and LDA”. *Information Retrieval* 14 (2): 178–203.

Lupton, Deborah. 2014. *Digital Sociology*. Abingdon, Oxon: Routledge.

Macaulay, Stewart. 1963. “Non-Contractual Relations in Business: A Preliminary Study”. *American Sociological Review* 28, no. 1 (): 55–67.

MacCormack, Alan, John Rusnak, and Carliss Y. Baldwin. 2006. “Exploring the Structure of Complex Software Designs: An Empirical Study of Open Source and Proprietary Code”. *Management Science* 52 (7): 1015–1030.

Macneil, Ian R. 1978. “Contracts: Adjustment of Long-Term Economic Relations under Classical, Neoclassical, and Relational Contract Law”. *Northwestern University Law Review* 72 (6): 854.

March, James G. 1995. “The Future, Disposable Organizations and the Rigidities of Imagination”. *Organization* 2 (3): 427–440.

Martin, John Levi. 2009. *Social structures*. Princeton: Princeton University Press.

Marx, Karl. 1992. *Capital: Volume 1: A Critique of Political Economy*. Reprint edition. Trans. by Ben Fowkes. London ; New York, N.Y: Penguin Classics.

Massa, Felipe G., and Siobhan O’Mahony. 2021. “Order from Chaos: How Networked Activists Self-Organize by Creating a Participation Architecture”. *Administrative Science Quarterly* ().

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

May, Timothy C. 1992. “The crypto anarchist manifesto”. Cypherpunk listserv. https://www.activism.net/cypherpunk/crypto-anarchy.html.

McMillan Cottom, Tressie. 2017. “Black Cyberfeminism: Ways forward for Intersectionality and Digital Sociology”. In *Digital Sociologies*, Reprint edition, ed. by Jessie Daniels, Karen Gregory, and Tressie McMillan Cottom, 211–231. Bristol, UK Chicago, IL: Policy Press.

Mejias, Ulises A., and Nick Couldry. 2019. “Datafication”. *Internet Policy Review* 8 (4). ISSN: 2197-6775, visited on 03/23/2021. https://policyreview.info/concepts/datafication.

Merkle, Ralph. 2016. “DAOS, Democracy and Governance”. *Cryonics Magazine* 37 (4): 28–40.

Merton, Robert K. 1942. “A Note on Science and Democracy”. *Journal of Legal and Political Sociology* 1-2:115–126.

Meyer, Alan D., Anne S. Tsui, and C. R. Hinings. 1993. “Configurational approaches to organizational analysis”. *Academy of Management Journal* 36 (6): 1175–1195.

Meyer, John W., and Brian Rowan. 1977. “Institutionalized Organizations: Formal Structure as Myth and Ceremony”. *American Journal of Sociology* 83 (2): 340–363.

Michels, Robert. 1966. *Political parties; a sociological study of the oligarchical tendencies of modern democracy*. Trans. by Eden Paul and Cedar Paul. New York: Free Press.

Miers, Ian, et al. 2013. “Zerocoin: anonymous distributed e-cash from Bitcoin”. Berkeley, CA: IEEE.

Mimno, David, et al. 2011. “Optimizing Semantic Coherence in Topic Models”. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 262–272. EMNLP ’11. Stroudsburg, PA, USA: Association for Computational Linguistics.

Mohr, John W., and Vincent Duquenne. 1997. “The Duality of Culture and Practice: Poverty Relief in New York City, 1888-1917”. *Theory and Society* 26 (2): 305–356.

Mollick, Ethan. 2014. “The dynamics of crowdfunding: an exploratory study”. *Journal of Business Venturing* 29 (1): 1–16.

Mumford, Lewis. 1964. “Authoritarian and Democratic Technics”. *Technology and Culture* 5 (1): 1–8. Visited on 10/10/2016.

Nagy, Peter, and Gina Neff. 2015. “Imagined affordance: reconstructing a keyword for communication theory”. *Social Media + Society* 1 (2): 1–9.

Nakamoto, Satoshi. 2008. *Bitcoin: a peer-to-peer electronic cash system*. Satoshi Nakamoto Institute. http://nakamotoinstitute.org/bitcoin/.

Narayanan, Arvind, and Jeremy Clark. 2017. “Bitcoin’s Academic Pedigree”. *Commun. ACM* 60 (12): 36–45.

Narayanan, Arvind, et al. 2016. *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction*. Princeton: Princeton University Press.

Navis, Chad, and Mary Ann Glynn. 2010. “How new market categories emerge: temporal dynamics of legitimacy, identity, and entrepreneurship in satellite radio, 1990-2005”. *Administrative Science Quarterly* 55 (3): 439–471.

Neff, Gina, and David Stark. 2004. “Permanently beta: responsive organization in the Internet era”. In *Society Online: The Internet in Context*, ed. by Philip N. Howard and Steve Jones, 173–188. Thousand Oaks, CA: SAGE Publications, Inc.

Nelson, Laura K. 2017. “Computational Grounded Theory: A Methodological Framework”. *Sociological Methods & Research*.

Nickerson, Jack A., and Todd R. Zenger. 2004. “A knowledge-based theory of the firm - The problem-solving perspective”. *Organization Science* 15 (6): 617–632.

Nisan, Noam, et al., eds. 2007. *Algorithmic Game Theory*. Cambridge, UK: Cambridge University Press.

North, Douglass C., and Barry R. Weingast. 1989. “Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England”. *The Journal of Economic History* 49 (4): 803–832. http://journals.cambridge.org/article_S0022050700009451.

O’Mahony, Siobhán. 2003. “Guarding the commons: how community managed software projects protect their work”. *Research Policy*, Open Source Software Development, 32 (7): 1179–1198.

– . 2002. “The emergence of a new commercial actor: community managed software projects”. PhD thesis, Stanford University.

O’Mahony, Siobhán, and Beth A. Bechky. 2008. “Boundary Organizations: Enabling Collaboration among Unexpected Allies”. *Administrative Science Quarterly* 53 (3): 422–459.

O’Mahony, Siobhán, and Fabrizio Ferraro. 2007. “The Emergence of Governance in an Open Source Community”. *Academy of Management Journal* 50, no. 5 (): 1079–1106.

O’Mahony, Siobhán, and Karim R. Lakhani. 2011. “Organizations in the shadow of communities”. In *Communities and Organizations*, ed. by Christopher Marquis, Michael Lounsbury, and Royston Greenwood, 33:3–36. Research in the Sociology of Organizations. Bingley, UK: Emerald Group Publishing Limited.

O’Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishers.

O’Mahony, Siobhán. 2007. “The governance of open source initiatives: what does it mean to be community managed?” *Journal of Management & Governance* 11 (2): 139–150.

Obstfeld, David. 2017. *Getting new things done: networks, brokerage, and the assembly of innovative action*. Stanford, CA: Stanford Business Books.

Of England, Bank. 2020. *Central Bank Digital Currency: Opportunities, Challenges and Design*. Discussion Paper. London, England: Bank of England. Visited on 04/07/2021. http://www.bankofengland.co.uk/paper/2020/central-bank-digital-currency-opportunities-challenges-and-design-discussion-paper.

Olson, Mancur. 1971. *The Logic of Collective Action: Public Goods and the Theory of Groups*. 2nd. Harvard University Press.

Orlikowski, Wanda J. 2002. “Knowing in Practice: Enacting a Collective Capability in Distributed Organizing”. *Organization Science* 13 (3): 249–273.

Orlikowski, Wanda J., and Susan V. Scott. 2008. “Sociomateriality: Challenging the Separation of Technology, Work and Organization”. *Academy of Management Annals* 2 (1): 433–474.

Orlikowski, Wanda J., and JoAnne Yates. 1994. “Genre Repertoire: The Structuring of Communicative Practices in Organizations”. *Administrative Science Quarterly* 39 (4): 541–574.

Ostrom, Elinor. 1990. *Governing the commons: the evolution of institutions for collective action*. Cambridge: Cambridge University Press.

– . 2003. “How Types of Goods and Property Rights Jointly Affect Collective Action”. *Journal of Theoretical Politics* 15 (3): 239–270.

– . 2017. “Polycentric Systems: Multilevel Governance Involving a Diversity of Organizations”. In *Global Environmental Commons: Analytical and Political Challenges in Building Governance Mechanisms*. Oxford Scholarship Online: Oxford University Press.

– . 2005. *Understanding Institutional Diversity*. Princeton: Princeton University Press.

Ostrom, Elinor, and Edella Schlager. 1996. “The Formation of Property Rights”. In *Rights to Nature*, ed. by Susan Hanna, Carl Folke, and Karl-Goran Maler, 127–156. Washington DC: Island Press.

Owen-Smith, Jason. 2001. “Managing Laboratory Work through Skepticism: Processes of Evaluation and Control”. *American Sociological Review* 66 (3): 427–452.

Parker, Geoffrey, Marshall Van Alstyne, and Xiaoyue Jiang. 2017. “Platform Ecosystems: How Developers Invert the Firm”. *MIS Quarterly* 41 (1): 255–A4.

Pasquale, Frank. 2019. “A Rule of Persons, Not Machines: The Limits of Legal Automation”. *George Washington Law Review* 87:1.

– . 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Perez, Carlota. 2003. *Technological Revolutions and Financial Capital: the Dynamics of Bubbles and Golden Ages*. Cheltenham: Edward Elgar Pub.

Philippon, Thomas. 2016. “The FinTech Opportunity”, National Bureau of Economic Research, 22476 ().

Pistor, Katherina. 2019. *The Code of Capital: How the Law Creates Wealth and Inequality*. Princeton: Princeton University Press.

Posner, Eric A., and E. Glen Weyl. 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press.

Powell, Walter W. 1990. “Neither Market Nor Hierarchy: Network Forms of Organization”. *Research in Organizational Behavior* 12:295–336.

Powell, Walter W., Kelley Packalen, and Kjersten Whittington. 2012. “Organizational and institutional genesis: the emergence of high-tech clusters in the life sciences”. In *The emergence of organizations and markets*, ed. by John Frederick Padgett and Walter W. Powell, 434–465. Princeton: Princeton University Press.

Powell, Walter W., and Kurt Sandholtz. 2012. “Chance, nécessité, et naïveté: ingredients to create a new organizational form”. In *The Emergence of Organizations and Markets*, ed. by John Frederick Padgett and Walter W. Powell, 379–433. Princeton: Princeton University Press.

Powell, Walter W., et al. 2017. “Institutional analysis in a digital era: mechanisms and methods to understanding emerging fields”. In *New Themes in Institutional Analysis: Topics and Issues from European Research*, ed. by C. Krucken et al., 305–344. Cheltenham, UK: Edward Elgar Publishing.

Prewitt, Matthew, and Steven McKie. 2018. “Blockchain Communities and Their Emergent Governance: Thoughts on Legitimacy and Narratives”. Medium. Visited on 03/09/2019. https://medium.com/amentum/blockchain-communities-and-their-emergent-governance-cfe5627dcf52.

Puranam, Phanish. 2018. *The Microstructure of Organizations*. Oxford: Oxford University Press.

Puranam, Phanish, Oliver Alexy, and Markus Reitzig. 2014. “What’s ”new” about new forms of organizing?” *Academy of Management Review* 39 (2): 162–180.

Rahwan, Iyad. 2018. “Society-in-the-loop: programming the algorithmic social contract”. *Ethics and Information Technology* 20 (1): 5–14.

Rao, Hayagreeva. 1998. “Caveat Emptor: The Construction of Nonprofit Consumer Watchdog Organizations”. *American Journal of Sociology* 103, no. 4 (): 912–961. ISSN: 0002-9602, visited on 01/19/2017. doi:10.1086/231293. http://www.journals.uchicago.edu/doi/abs/10.1086/231293.

Rawls, John. 2005. *Political Liberalism*. Expanded. Columbia University Press.

Resnick, Paul, and Hal R. Varian. 1997. “Recommender Systems”. *Commun. ACM* 40 (3): 56–58. ISSN: 0001-0782.

Resnick, Paul, and Richard Zeckhauser. 2002. “Trust among strangers in internet transactions: Empirical analysis of eBay’s reputation system”. In *The Economics of the Internet and E-commerce*, ed. by Michael R. Baye, 11:127–157. Advances in Applied Microeconomics. Emerald Group Publishing Limited.

Riksbank, Sveriges. 2021. *E-Krona Pilot: Phase 1*. Stockholm: Sveriges Riksbank.

Roberts, Margaret, et al. 2018. “stm: Estimation of the Structural Topic Model”. https://CRAN.R-project.org/package=stm.

Rogers, Everett M. 1995. *Diffusion of innovations*. 4th ed. New York: Free Press.

Rosenkopf, Lori, and Michael L. Tushman. 1998. “The Coevolution of Community Networks and Technology: Lessons from the Flight Simulation Industry”. *Industrial and Corporate Change* 7 (2): 311–346.

Roth, Alvin E. 2002. “The economist as engineer: game theory, experimentation and computation as tools for design economics”. *Econometrica* 70 (4): 1341–1378.

– . 2008. “What Have We Learned from Market Design?” *The Economic Journal* 118 (527): 285–310.

Roughgarden, Tim. 2010. “Algorithmic game theory”. *Communications of the ACM* 53 (7): 78–86.

Rozas, David, and Steven Huckle. 2020. “Loosen control without losing control: Formalization and decentralization within commons-based peer production”. *Journal of the Association for Information Science and Technology*: 1–20.

Ruef, Martin. 2000. “The emergence of organizational forms: a community ecology approach”. *American Journal of Sociology* 106 (3): 658–714.

Rumburg, Roneil, Sid Sethi, and Hareesh Nagaraj. 2019. *A Decentralized Protocol for Audio Content*. Visited on 09/25/2019. https://whitepaper.audius.co/AudiusWhitepaper.pdf.

Sabel, Charles F., and Jonathan Zeitlin. 1997. “Stories, Strategies, Structures: Rethinking Historical Alternatives to Mass Production”. In *World of Possibilities: Flexibility and Mass Production in Western Industrialization*, ed. by Charles F. Sabel and Jonathan Zeitlin, 1–33. Cambridge: Cambridge University Press.

Sadowski, Jathan. 2019. “When Data is Capital: Datafication, Accumulation, and Extraction”. Publisher: SAGE Publications Ltd, *Big Data & Society* 6 (1): 1–12. https://doi.org/10.1177/2053951718820549.

Sah, Raj Kumar, and Joseph E. Stiglitz. 1986. “The Architecture of Economic Systems: Hierarchies and Polyarchies”. *The American Economic Review* 76 (4): 716–727.

Sassen, Saskia. 1999. “Digital Networks and Power”. In *Spaces of Culture: City, Nation, World*, ed. by Mike Featherstone and John Lash, 49–63. London: SAGE Publications Ltd.

– . 2002. “Towards a sociology of information technology”. *Current Sociology* 50 (3): 365–388.

Schelling, Thomas C. 1981. *The Strategy of Conflict*. Reprint edition. Cambridge: Harvard University Press.

Schlager, Edella, and Elinor Ostrom. 1992. “Property Rights Regimes and Natural Resources: A Conceptual Analysis”. *Land Economics* 68 (3): 249–262.

Schneider, Fred B. 1990. “Implementing Fault-tolerant Services Using the State Machine Approach: A Tutorial”. *ACM Comput. Surv.* 22, no. 4 (): 299–319. ISSN: 0360-0300, visited on 12/03/2019. doi:10.1145/98163.98167. http://doi.acm.org/10.1145/98163.98167.

Schneier, Bruce. 1998. “Memo to the Amateur Cipher Designer”. Schneier on Security. https://www.schneier.com/crypto-gram/archives/1998/1015.html#cipherdesign.

Schor, Juliet B., and William Attwood-Charles. 2017. “The ”sharing” economy: labor, inequality, and social connection on for-profit platforms”. *Sociology Compass* 11 (8): e12493.

Seaver, Nick. 2017. “Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems”. *Big Data & Society* 4 (2): 1–12.

– . 2019. “Knowing Algorithms”. In *digitalSTS: A Field Guide for Science & Technology Studies*, ed. by Janet Vertesi and David Ribes, 412–422. Princeton: Princeton University Press.

Seidel, Marc-David L. 2018. “Questioning Centralized Organizations in a Time of Distributed Trust”. *Journal of Management Inquiry* 27 (1): 40–44.

Seidel, Marc-David L., and Henrich R. Greve. 2017. “Emergence: How Novelty, Growth and Formation Shape Organizations and Their Ecosystems”. In *Research in the Sociology of Organizations*, 50:1–27. Emerald Publishing Limited.

Shapiro, Susan P. 1987. “The Social Control of Impersonal Trust”. *American Journal of Sociology* 93 (3): 623–658.

Shaw, Aaron, and Benjamin M. Hill. 2014. “Laboratories of Oligarchy? How the Iron Law Extends to Peer Production”. *Journal of Communication* 64 (2): 215–238.

Star, Susan Leigh, and James R. Griesemer. 1989. “Institutional ecology, ’translations’ and boundary objects: amateurs and professionals in Berkeley’s Museum of Vertebrate Zoology, 1907-39”. *Social Studies of Science* 19 (3): 387–420.

Star, Susan Leigh, and Karen Ruhleder. 1996. “Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces”. *Information Systems Research* 7 (1): 111–134.

Stark, David. 1996. “Recombinant Property in East European Capitalism”. *American Journal of Sociology* 101, no. 4 (): 993–1027.

– . 2009. *The sense of dissonance: accounts of worth in economic life*. Princeton: Princeton University Press.

Starr, Paul. 2019. *Entrenchment: Wealth, Power, and the Constitution of Democratic Societies*. New Haven, CT: Yale University Press.

– . 1992. “Social Categories and Claims in the Liberal State”. *Social Research* 59 (2): 263–295.

– . 2005. *The Creation of the Media: Political Origins of Modern Communications*. New York: Basic Books.

– . 2021. “The Relational Public”. *Sociological Theory*: 07352751211004660. Visited on 05/06/2021. https://doi.org/10.1177/07352751211004660.

– . 1982. *The Social Transformation of American Medicine*. Basic Books.

Strang, David, and John W. Meyer. 1993. “Institutional Conditions for Diffusion”. *Theory and Society* 22 (4): 487–511.

Sugden, Robert. 1995. “A Theory of Focal Points”. *The Economic Journal* 105 (430): 533–550.

Swartz, Lana. 2017. “Blockchain Dreams: Imagining Techno-Economic Alternatives after Bitcoin”. In *Another Economy is Possible: Culture and Economy in a Time of Crisis*, ed. by Manuel Castells, 82–105. Google-Books-ID: DjFKDwAAQBAJ. John Wiley & Sons.

– . 2018. “What was Bitcoin, what will it be? The techno-economic imaginaries of a new money technology”. *Cultural Studies* 32 (4): 623–650.

Swedberg, Richard. 2000. “The Social Science View of Entrepreneurship: Introduction and Practical Applications”. In *Entrepreneurship: The Social Science View*, ed. by Richard Swedberg, 7–44. Oxford: Oxford University Press.

Swidler, Ann. 1986. “Culture in Action: Symbols and Strategies”. *American Sociological Review* 51 (2): 273–286.

Sydow, Jörg, Lars Lindkvist, and Robert DeFillippi. 2004. “Project-Based Organizations, Embeddedness and Repositories of Knowledge: Editorial”. *Organization Studies* 25 (9): 1475–1489. ISSN: 0170-8406.

Sydow, Jörg, and Udo Staber. 2002. “The Institutional Embeddedness of Project Networks: The Case of Content Production in German Television”. *Regional Studies* 36, no. 3 (): 215–227.

Szabo, Nick. 1997. “Formalizing and Securing Relationships on Public Networks”. *First Monday* 2 (9). Visited on 09/06/2016. http://firstmonday.org/ojs/index.php/fm/article/view/548/469%23Contracts.

– . 1998. “Secure Property Titles with Owner Authority”. Satoshi Nakamoto Institute. Visited on 02/15/2017. http://nakamotoinstitute.org/secure-property-titles/#selection-24.0-24.10.

Tapscott, Don, and Alex Tapscott. 2016. *Blockchain revolution: how the technology behind Bitcoin Is changing money, business, and the world*. New York: Portfolio.

Tavory, Iddo, and Stefan Timmermans. 2014. *Abductive Analysis: Theorizing Qualitative Research*. Chicago: The University of Chicago Press.

Teece, David J. 2010. “Business Models, Business Strategy and Innovation”. *Long Range Planning*, Business Models, 43 (2): 172–194.

– . 2018. “Profiting from innovation in the digital economy: Enabling technologies, standards, and licensing models in the wireless world”. *Research Policy* 47 (8): 1367–1387.

Thompson, James D. 1967. *Organizations in action: Social science bases of administrative theory*. New York: McGraw-Hill.

Trajtenberg, Manuel. 2018. *AI as the next GPT: a Political-Economy Perspective* w24245. National Bureau of Economic Research. https://www.nber.org/papers/w24245.

Tushman, Michael L., and Philip Anderson. 1986. “Technological discontinuities and organizational environments”. *Administrative Science Quarterly* 31 (3): 439–465.

Tushman, Michael L., and Lori Rosenkopf. 1992. “Organizational determinants of technological change: toward a sociology of technological evolution”. In *Research in Organizational Behavior*, 14:311–347. JAI Press.

Unger, Roberto Mangabeira. 1987. *False Necessity: Anti-Necessitarian Social Theory in the Service of Radical Democracy*. Cambridge, UK: Cambridge University Press.

Van Lente, Harro, and Arie Rip. 1998. “The Rise of Membrane Technology: From Rhetorics to Social Reality”. *Social Studies of Science* 28, no. 2 (): 221–254.

Van Steen, Maarten, and Andrew S. Tanenbaum. 2017. *Distributed Systems*. 3.01 edition. The Netherlands: CreateSpace Independent Publishing Platform.

Vaughan, Diane. 2014. “Analogy, cases and comparative social organization”. In *Theorizing in social science: The context of discovery*, ed. by Richard Swedberg. Stanford: Stanford University Press.

– . 1998. “Rational choice, situated action, and the social control of organizations”. *Law & Society Review* 32 (1): 23–61.

Vigna, Paul, and Michael J. Casey. 2016. *The Age of Cryptocurrency: How Bitcoin and the Blockchain Are Challenging the Global Economic Order*. Reprint edition. New York: Picador.

Von Hippel, Eric, and Georg von Krogh. 2003. “Open Source Software and the “Private-Collective” Innovation Model: Issues for Organization Science”. *Organization Science* 14 (2): 209–223.

Wang, Licheng, et al. 2019. “Cryptographic Primitives in Blockchains”. *Journal of Network and Computer Applications*, no. 127: 43–58.

Wareham, Jonathan, Paul B. Fox, and Josep Lluís Cano Giner. 2014. “Technology Ecosystem Governance”. *Organization Science* 25 (4): 1195–1215.

Warren, Will. 2018. “Governance in 0x Protocol”. Medium. Visited on 03/04/2020. https://blog.0xproject.com/governance-in-0x-protocol-86779ae5809e.

Watkins, Elizabeth Anne, and David Stark. 2018. "The Möbius Organizational Form: Make, Buy, Cooperate, or Co-opt?" *Sociologica* 12 (1): 65–80. ISSN: 1971-8853, visited on 06/17/2020.

Weber, Max. 1978. *Economy and Society: An Outline of Interpretive Sociology*. Guenther Roth and Claus Wittich, eds. University of California Press.

Weick, Karl E. 1995. *Sensemaking in organizations*. Thousand Oaks, CA: SAGE.

– . 1990. "Technology as equivoque: sensemaking in new technologies". In *Technology and Organizations*, ed. by P.S. Goodman and L. Sproull, 1–44. San Francisco, CA: Jossey Bass.

– . 1993. "The collapse of sensemaking in organizations: the Mann Gulch Disaster". *Administrative Science Quarterly* 38 (4): 628–652.

Weiser, Mark. 1993. "Ubiquitous Computing". *Computer* 26:71–72. doi:10.1109/2.237456.

Werbach, Kevin. 2018. *The blockchain and the new architecture of trust*. Cambridge: The MIT Press.

West, Joel, and Siobhán O'Mahony. 2008. "The Role of Participation Architecture in Growing Sponsored Open Source Communities". *Industry and Innovation* 15 (2): 145–168.

Wheat, Rachel E., et al. 2013. "Raising money for scientific research through crowdfunding". *Trends in Ecology & Evolution* 28 (2): 71–72. Visited on 07/24/2020.

Winner, Langdon. 1996. "Who will we be in cyberspace?" *The Information Society* 12 (1): 63–72.

Wood, Gavin. 2014. *Ethereum: A Secure Decentralized Transaction Ledger*. Ethereum Yellow Paper.

Woodward, Joan. 1958. "Management and technology". In *Organizational Theory: Selected Readings*, Second, ed. by D.S. Pugh. New York, NY: Penguin Press.

Wry, Tyler, Michael Lounsbury, and Mary Ann Glynn. 2011. "Legitimating nascent collective identities: coordinating cultural entrepreneurship". *Organization Science* 22 (2): 449–463.

Xie, Pengtao, and Eric P. Xing. 2013. "Integrating Document Clustering and Topic Modeling". *arXiv:1309.6874 [cs, stat]*. arXiv: 1309.6874.

Yang, Tiantian, and Howard E. Aldrich. 2012. “Out of sight but not out of mind: why failure to account for left truncation biases research on failure rates”. *Journal of Business Venturing* 27 (4): 477–492.

Yates, Joanne, and Wanda J. Orlikowski. 1992. “Genres of Organizational Communication: A Structurational Approach to Studying Communication and Media”. *Academy of Management Review* 17 (2): 299–326.

Yin, Robert K. 2009. *Case Study Research: Design and Methods*. Vol. 4th ed. Applied social research methods series ; v. 5. Los Angeles: Sage Publications.

Zamfir, Vlad. 2018. “Blockchain Governance 101”. Good Audience. Visited on 03/09/2019. https://blog.goodaudience.com/blockchain-governance-101-eea5201d7992.

Zelizer, Viviana A. 2011. *Economic lives: how culture shapes the economy*. Princeton University Press.

– . 1979. *Morals and Markets: The Development of Life Insurance in the United States*. New York: Columbia University Press.

– . 1997. *The Social Meaning of Money*. Princeton University Press.

Zittrain, Jonathan. 2008. *The Future of the Internet and how to Stop It*. New Haven: Yale University Press.

Zott, Christoph, Raphael Amit, and Lorenzo Massa. 2011. “The Business Model: Recent Developments and Future Research”. *Journal of Management* 37 (4): 1019–1042.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

Zucker, Lynne G. 1986. “Production of trust: institutional sources of economic structure, 1840-1920”. *Research in Organizational Behavior* 8:53–111.

Zuckerman, Ezra W. 2017. “The Categorical Imperative Revisited: Implications of Categorization as a Theoretical Tool”. In *From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads*, 51:31–68. Research in the Sociology of Organizations. Emerald Publishing Limited.

– . 1999. “The categorical imperative: securities analysts and the illegitimacy discount”. *American Journal of Sociology* 104 (5): 1398–1438.