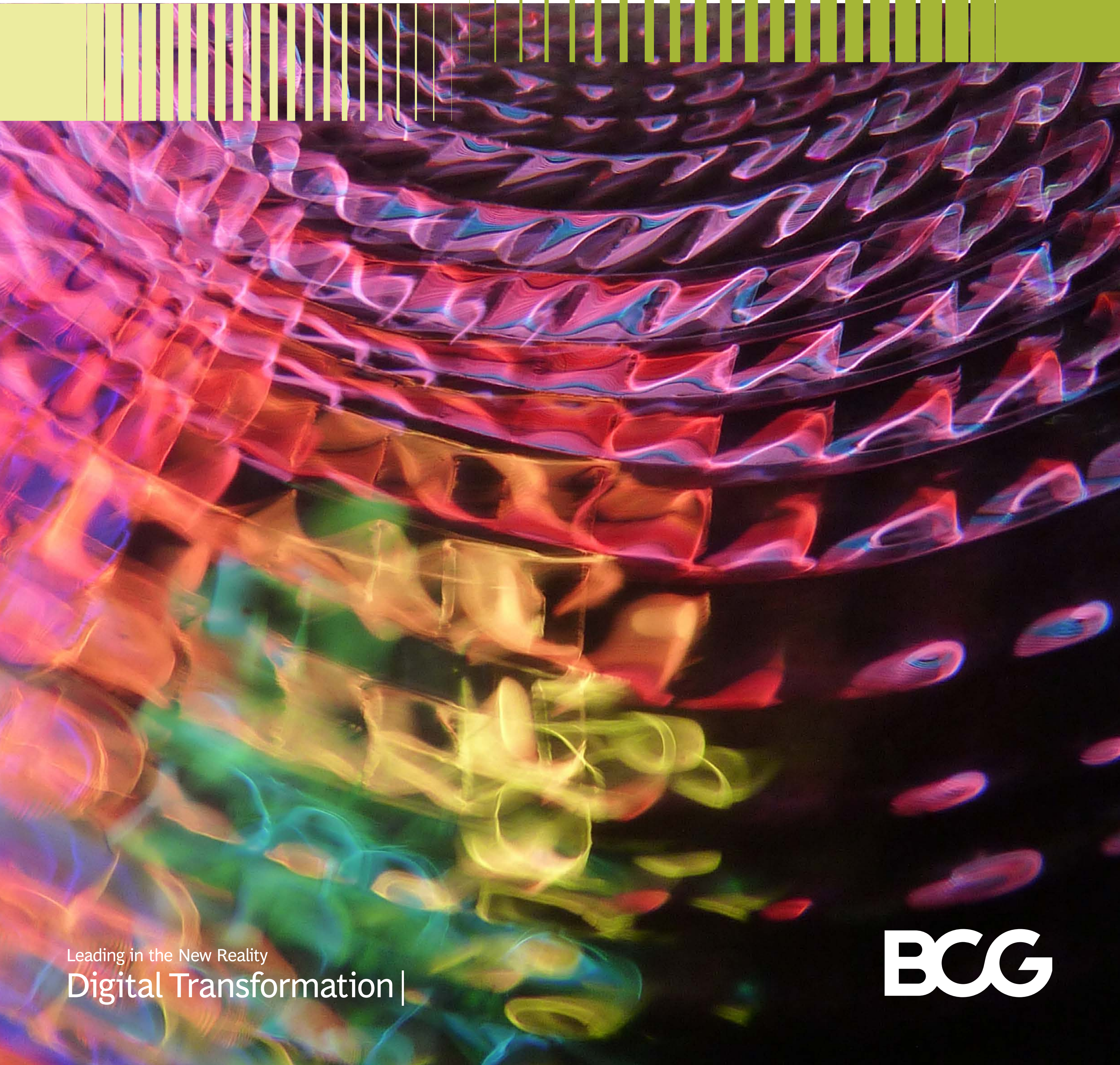
Beyond Blockchain

The Promise of Digital Trust Networks

December 2021
By Philip Evans, Marcos Aguiar, Matt Williams, and Santino Lacanna

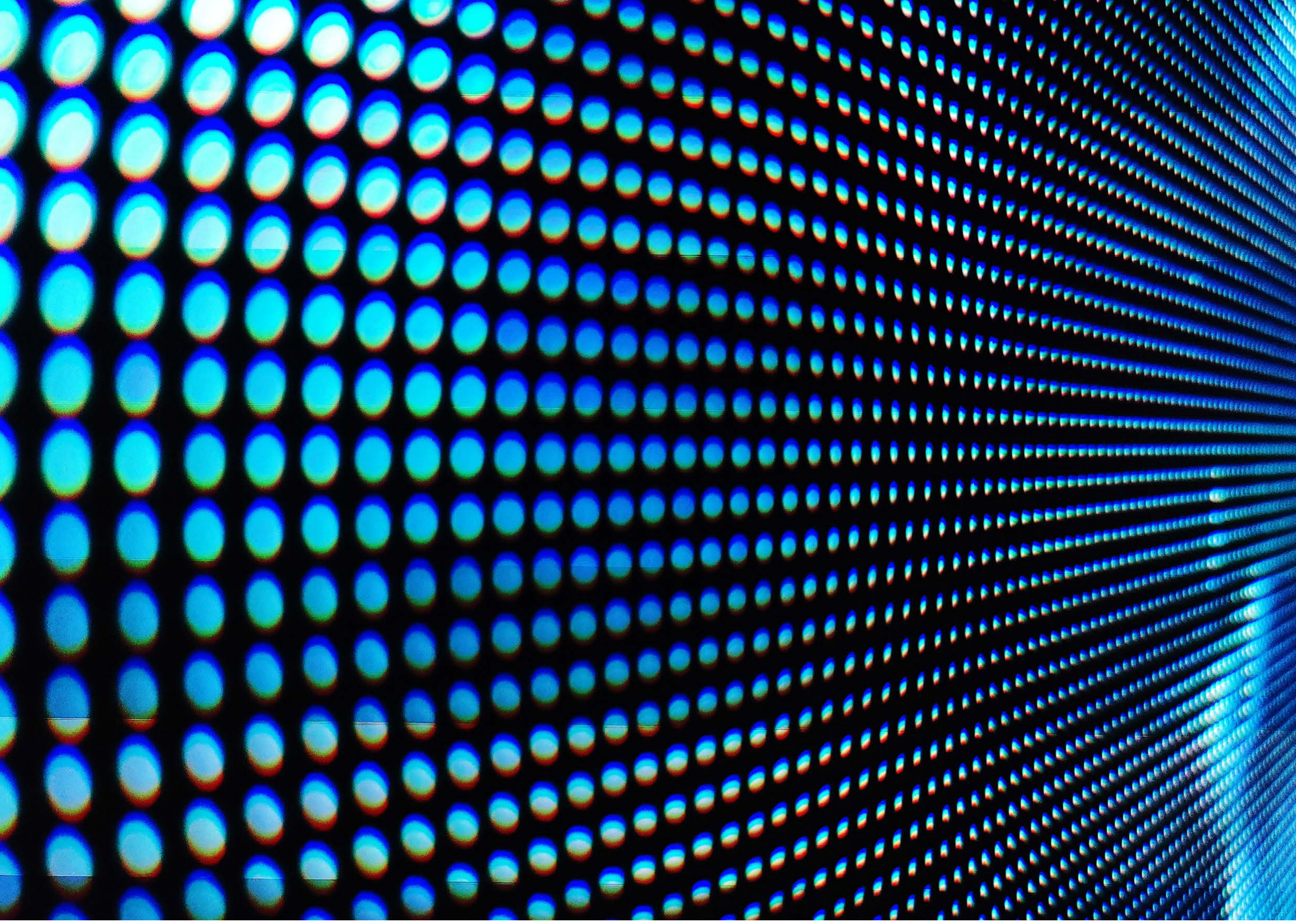


Leading in the New Reality
Digital Transformation |

Boston Consulting Group partners with leaders in business and society to tackle their most important challenges and capture their greatest opportunities. BCG was the pioneer in business strategy when it was founded in 1963. Today, we work closely with clients to embrace a transformational approach aimed at benefiting all stakeholders—empowering organizations to grow, build sustainable competitive advantage, and drive positive societal impact.

Our diverse, global teams bring deep industry and functional expertise and a range of perspectives that question the status quo and spark change. BCG delivers solutions through leading-edge management consulting, technology and design, and corporate and digital ventures. We work in a uniquely collaborative model across the firm and throughout all levels of the client organization, fueled by the goal of helping our clients thrive and enabling them to make the world a better place.

Contents



Introduction

The current excitement about blockchain is more than a reflection of speculative excess; in the business world, it is also a reflection of the intense interest in what some call the "architecture" of trust—how trust is designed into counterparty relationships.[1]

At a time when digital interactions are rapidly supplanting physical ones in business and civic life, that intensifying interest makes sense: trust is a precious enabler of transactions and collaboration, both internal and external. It is also expensive. Its economics are increasingly shaping the boundaries of the corporation—including, notably, the rise of the ecosystem in modern economies. Yet trust has proved largely immune to technical progress. It is complex, engendered by multiple elements, and works (and scales) in different ways.

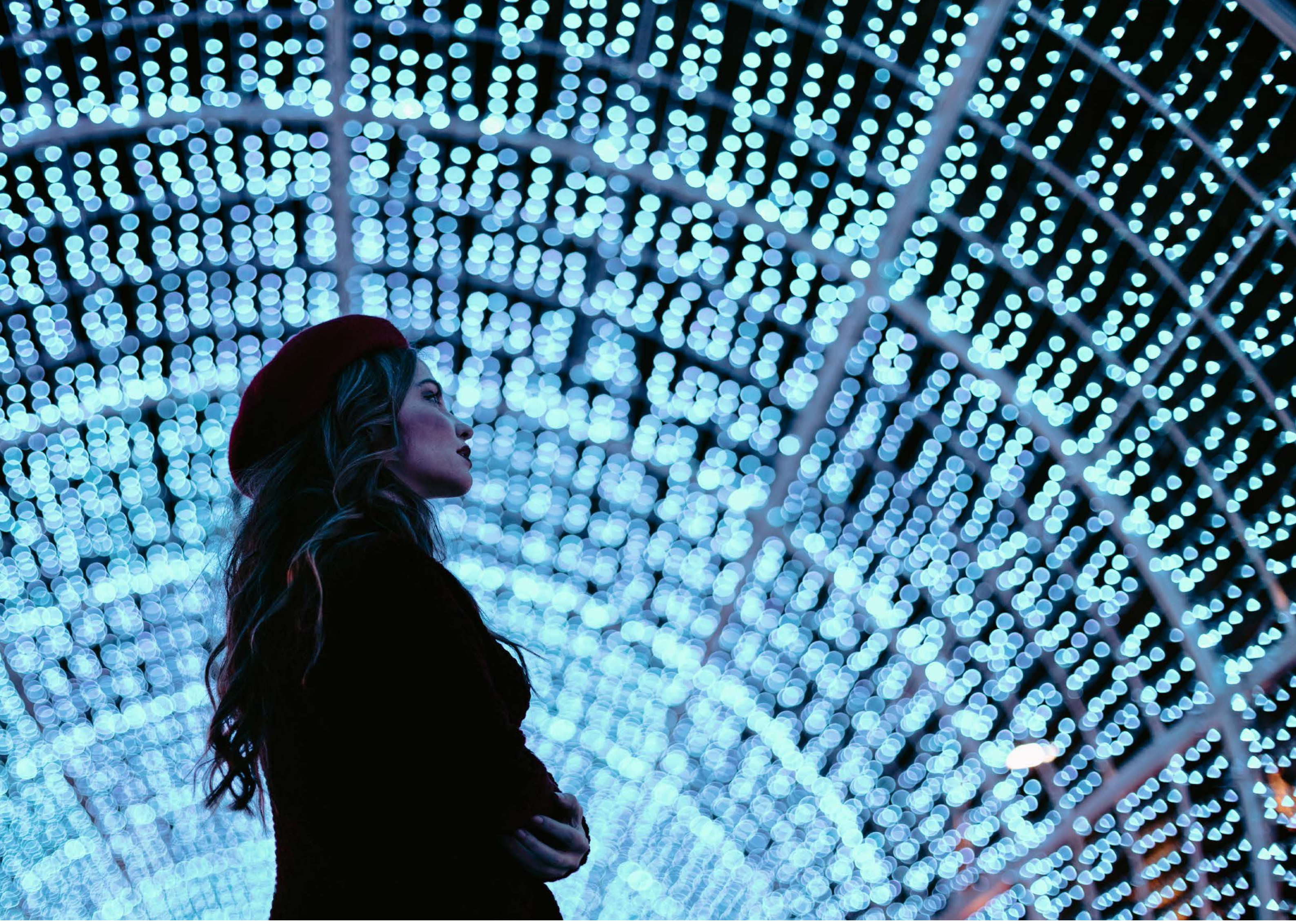
[1] This is well represented in the title of Kevin Werbach's excellent book *The Blockchain and the New Architecture of Trust* (MIT Press, 2018).

Blockchains—shared, distributed, and irrevocable ledgers —are a much-discussed solution. But many conflate blockchains with digital trust, when in fact blockchains deal with just one aspect of that architecture: the need for reliable ***intermediary*** recordkeeping. There is a second, much larger, issue that blockchains do not fully address: the inherent mutual distrust among ***counterparties***. Another set of mechanisms is required to address that issue, enabling what we call a digital trust network (DTN). A DTN is a set of digital mechanisms that facilitates the generation of mutual trust (or the management of mutual distrust) among transacting counterparties. Blockchain is one such mechanism—but only one.

This paper addresses the rationale underlying DTNs, clarifying their important distinctions from a blockchain. It also addresses their significant potential. To do so, we must go back to first principles: to understand the various forms of trust, identify the mechanisms by which trust and distrust are managed, and then trace the current and future impact of digital technology.

Trust and distrust have been addressed within many intellectual disciplines that often talk past each other. Game theorists, economists, psychologists, philosophers, and sociologists have severally provided powerful insights, each from a unique methodological perch. Moreover, once one recognizes that there are different bases of trust, it is a trivial insight to the business strategist that those bases must coexist in segments whose boundaries are defined by competitive advantage. Factors such as strength and scale economies define those relative boundaries. And more important, technology is transforming them.

In this paper, we provide an integrative, top-down view of DTNs. We focus on five bases of trust that serve as the foundation for the mechanisms that come together in a stack to form the DTN. We explore the mechanisms that engender trust and those that manage distrust—and the complex dynamics of both. The impact of technology on trust is examined through its direct impact on these mechanisms. They, in turn, influence the scale and power of the bases of trust, ultimately enabling trust among counterparties.



Why Digital Trust Networks?

In transactions, intermediaries are generally in the business of being trusted: for example, it is simply not in a bank’s business interests to misplace a payment. But there are plenty of cases, especially in less mature economies, where the trustworthiness of intermediaries remains an issue. The shared, irrevocable, distributed ledger of a blockchain can address this challenge. And intermediaries sometimes extract monopoly rents, so a blockchain—owned by everybody and by nobody—is a means for existing counterparties to update their transactions technology without ceding economic hegemony to a third party.

A blockchain may serve as irrevocable proof that a counterparty made a claim or promise in the database, but it does not prove that the claim in question is true or that the promise will be honored in the world at large. A blockchain provides an irrefutable account of its own history, not of the world. Blockchains can be extended to partially address the issue of counterparty distrust via end-to-end digitization of a transactions process—through the Internet of Things (IoT), data standards, “oracles,” and the automated execution of complex instructions in “smart contracts," all connecting to a common logical database.[2] This process is as trustworthy as the real-world entities providing inputs. Thus, escrow services can be executed on a blockchain such as Ethereum, possibly enabling greater efficiency and greater counterparty trust. However, blockchain per se is not adding to efficiency (indeed the technology is inherently inefficient). It is adding to intermediary trust, if indeed that is a problem.

A DTN serves a broader role than a blockchain. It is a set of end-to-end digital protocols and processes that use a common virtual database (among other things) to enable secure and trusted transactions among counterparties. It is a system that structures interactions through standardized interfaces and possibly dedicated intermediaries in order to engender trust or manage distrust among transacting parties. It will generally require a recordkeeping system that may or may not be maintained on a blockchain.

DTNs come in many varieties and are generally embedded within digital ecosystems that serve wider purposes; for example:

- Uber, and similarly Airbnb and Amazon Marketplace, engendering trust and managing distrust among buyers and sellers
- The Apple iOS ecosystem, in which Apple curates the content and controls the behavior of developers and users via code and policies
- So-called "zero trust" security systems implemented within corporations to control the behavior of and access by employees and outsiders
- India's Aadhaar architecture, which enables a range of trust-dependent services on a secure foundation of digital identity for the country's nearly 1.4 billion citizens
- So-called decentralized autonomous organizations (DAOs): mechanisms by which investors achieve some of the functionality of a corporation via smart contracts executed on a blockchain

Blockchains and DTNs are conceptually distinct: a blockchain is a database; a DTN is a process architecture. Blockchains may run as the data hub for a DTN, but you can have a blockchain without a DTN (Bitcoin), or a DTN without a blockchain (Uber). If you stand up an Uber driver, you still get charged and they still get paid; that is a smart contract (effectively escrow) but one that is programmed and executed digitally on the Uber system and Uber's cloud database, not on a blockchain. Of course, this arrangement requires that both passenger and driver trust Uber, but that is part of its business model. Blockchain and DTN are thus entirely separate and separable technologies: the first, a ledger; the second, a set of executables and digital interfaces; the first addressing the (less common) problem of distrust of ***intermediaries***, and the second addressing the (pervasive) problem of distrust among ***counterparties***.

[2] There is no requirement that this common logical database be in one location or under the control of a single institution (a blockchain is neither).

Trust

What It Is, How It Works

"Trust" in everyday speech is a fuzzy concept, and for our purposes we need to sharpen it a bit. Trust researchers generally define the term to mean *the willingness of one party to be vulnerable to the actions of another*.[3] The operative word here is "party"; that is, a rational, intentional agent. People are parties, of course, but so are corporations to the extent that we think of them as having goals and making reasoned choices.

This excludes certain kinds of trust. We may speak loosely of "trusting a blockchain" or "trusting AI," but because neither blockchains nor AI are rational, intentional agents, what we really mean in those cases is "relying on" not "trusting." Trust, as we use the term, always requires reliance, but reliance does not require trust. Some writers, however, define trust to include reliance. This allows, for example, a transactor to "trust" a blockchain even though it is obviously not a rational, intentional agent. Concepts

[3] This is a close paraphrase of "the willingness of a party to be vulnerable to the actions of another party," the definition used by R. C. Mayer, J. Davis, and F. Schoorman, "An Integrative Model of Oganizational Trust," *Academy of Management Review*, 20 (1995): 709–34. This definition is cited approvingly by F. Lumineau, O. Schilke, and W. Wang: "Organizational Trust in the Age of the Fourth Industrial Revolution: Shifts in the Nature, Production, and Targets of Trust" (currently under journal review). It is also essentially identical to the definition ("the decision to rely on another party [i.e., person, group, or organization] under a condition of risk") employed by S. C. Currall and A. C. Inkpen, "On the Complexity of Organizational Trust: A Multi-Level Co-Evolutionary Perspective and Guidelines for Future Research," in *Handbook of Trust Research*, ed. R. Bachmann and A. Zaheer (Edward Elgar Publishing, 2006), chapter 13.

such as Niklas Luhmann's "systemic trust" (***Systemvertrauen***) employ this broader definition. In our parlance, parties making, say, a Bitcoin payment ***rely*** on the ledger but they do not ***trust*** it: such a transaction is "trustless." The mechanisms for managing the absence of trust are discussed later in this paper.

Secondly, in common parlance, an institution may be trusted (or not) outside the context of any specific transaction. Companies such as Uber and Exxon have been the objects of backlash by customers appalled by their various managerial practices. This is sometimes characterized as a "loss of trust" in these companies. And sociologists conduct regular surveys of people's trust in institutions in general. But this is not an issue of trust in the sense that we are employing. We are discussing what might be called "transactional trust"; this "generalized trust" is a different—but no less important—topic.[4]

Trust is also relative to a particular transaction, context, or vulnerability. We might trust a nanny with the keys to the house but not the combination to the safe, or trust an eBay vendor with our shipping address but not our bank account information.

> **The trust challenge is twofold: how to generate trust among parties, and how to manage transactions to minimize their need for trust, or substitute for trust.**

Technology, as we will argue, is transforming both.

Two Fundamental Principles

Trust generation can be grounded in either of two fundamental principles, which might be called "reciprocity" and "perceived trustworthiness."

- *Reciprocity* is the willingness of the trustor (the one doing the trusting) to make themselves vulnerable to the trustee (the one being—or not being—trusted) in the belief that the trustee sees it in their own *rational self-interest* to take the trustor's specific interests into account. The trustor's interests are contained, or "encapsulated," in the interests of the trustee. The purest case of reciprocity is found in game-theory experiments, where anonymous players of repeated prisoner's dilemma games generally learn to play collaboratively (trust each other), because there is greater value in sustained future cooperation than in winning any one game.[5]
- *Perceived trustworthiness* is the trustor's observation that the trustee is motivated by morals or social norms to behave in a principled manner, warranting trust. The trustor imputes values or goals to the trustee. Perceived trustworthiness is based on observation of the trustee in contexts outside the immediate transaction and/or by the trustee's membership or participation in groups that (the trustor observes) are characterized by such values or goals.

Reciprocity can be described as the "shadow of the future" because trust today is predicated on the return of favors tomorrow. Trustworthiness, on the other hand, is the "shadow of the past" because it is inferred from history. These two principles are embedded in the bases on which a party may decide to trust.[6]

The Five Bases of Trust

There are five bases of trust: Perceived Reputation-at-Risk, Attributed Norms, Empathy, Shared Identity, and Relationship. (See Exhibit 1.)

These bases are all cognitive; that is, they are perceptions, attributions, or sympathies in the mind of the trustor. As mental states, they are invisible to the outside observer. They vary in ***strength*** (How much vulnerability will the trustor accept?) and also in ***scalability*** (How much effort is required to establish trust on the given basis—and therefore, with how many parties can the trustor sustain trust?)

> **Some types of trust, such as trust in a consumer brand, are cheap to establish and can thus be done at scale. Others, such as relationship trust, are powerful but costly, requiring one-on-one time and commitment (for example, marital trust).**

Strength and scalability, therefore, vary inversely.

Let's look at the bases of trust in roughly ascending order of strength and descending order of scalability.

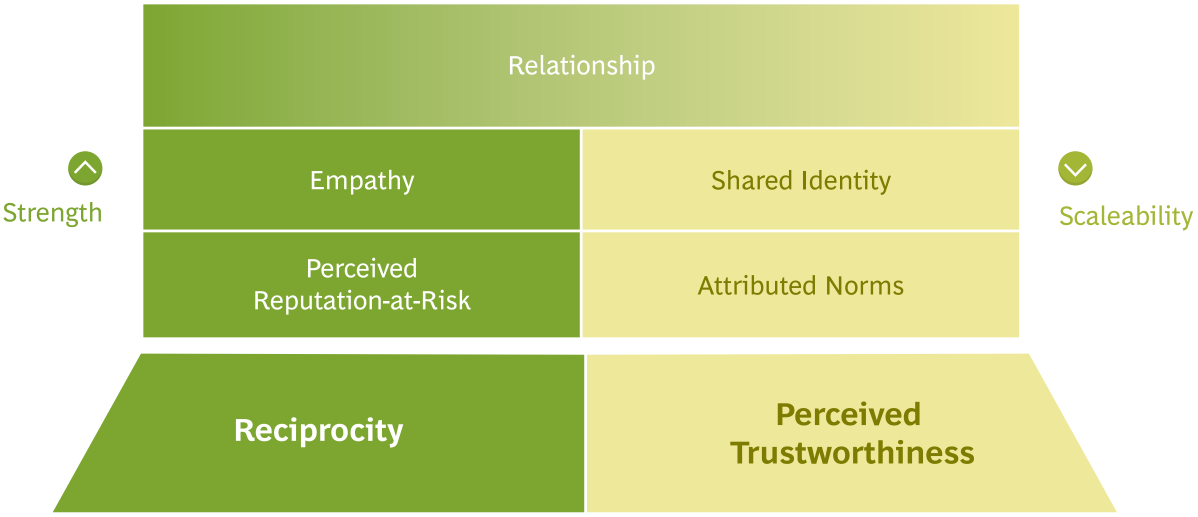
Perceived Reputation-at-Risk. This basis of trust requires a third party: the "audience" that will observe the current transaction and make inferences about the trustee's behavior. They will act on those inferences in future contexts; knowing this, the trustee is rationally motivated to present themselves as reliable to future potential counterparties. The current trustor, aware of this motivation, therefore trusts the trustee. Reputation-at-Risk is thus an asset that is put up as collateral in the transaction; and the larger the relevant audience (parties with whom the

[4] The crisis of generalized trust, and how to deal with it, are eloquently described in S. Sucher and S. Gupta, *The Power of Trust* (Public Affairs, July 2021).

[5] R. Axelrod and W. D. Hamilton, 'The Evolution of Cooperation," *Science* 211, no. 4489 (March 1981): 1390–96. In the academic literature, the term "encapsulation" is often preferred; this is owing to R. Hardin, *Trust and Trustworthiness* (Russell Sage Foundation, 2002), chapter 1.

[6] In this paper, we avoid discussion of "generalized trust" and "social capital" (the latter, another idea prominent in political sociology). Both are best understood as combinations of the bases we do discuss, and different scholars have used the words differently.

Exhibit 1 - The Five Bases of Trust and Their Underlying Principles



The two principles supporting the bases of trust

Source: BCG Henderson Institute analysis.

trustee might have future transactions), the greater the strength of the trust so engendered.

Attributed Norms. The trustor trusts based on the direct or indirect observation of the trustee's behavior as a member of a group characterized by particular norms or values. In principle, that trust can be established in three ways: by directly observing the trustee's behavior (independent of any group context); by relying on the testimony of others (Reputation-by-Hearsay); or by inferring the norms and values of a group and trusting someone—otherwise unknown—purely on the basis of their membership in it.[7] (Thus, we may trust a clergyman or an "officer and a gentleman," even without knowing them very well, by merit of the values we attribute to the clergy or the officer corps.) In practice, these three means are almost always deeply intertwined: the trustor has some direct observation of the trustee in other contexts, the trustor learns from others who know the trustee, and the trustor has a perception of the characteristic trustworthiness of a group and thus trusts those "others" by virtue of their membership of that same group. These bases are mutually reinforcing.[8]

Both Reputation-at-Risk and Attributed Norms generally require that the trustor observe the trustee's participation in a (typically large) group, but they work in different ways: the former through rational self-interest; the latter through imputed values.

Empathy. Humans have a universal, instinctual propensity to trust one another. Trustees signal their intentions by body language, smiling, eye contact, tone of voice, mimicry, and so forth. These are largely unconscious and involuntary—"honest signals" (in the vernacular of evolutionary biology) that communicate reliable information. Equally unconscious and involuntary is the favorable reception of those signals by trustors. Empathy evokes and conveys trust in both directions simultaneously.

[7] We are sharply distinguishing "Reputation-at-Risk," which is a form of reciprocity, from "Reputation-by-Hearsay," which is a form of perceived trustworthiness. Reputation-by-Hearsay is merely information about the trustee: it requires that some third parties have observed the trustee in other contexts and report to the trustor. The trustee does not need to know that the communication is taking place; the third parties do not need to observe the current transaction; the trustee does not need to know that it is being observed by third parties; and the trustor does not need to know that the trustee knows that they are observing it. All of these are requirements for Reputation to be "at risk." Reputation-at-Risk is far more powerful than Reputation-by-Hearsay, and as we will argue later, technology is enabling it to substitute.

[8] The importance of origins and group membership in anchoring trust is amply illustrated in literature. In Richard Wagner's *Lohengrin*, Elsa of Brabant is falsely accused of murder. A knight in (literally) shining armor appears on a boat drawn by a swan and offers to defend her honor in single combat. He asks her to swear that she will never ask his name, heritage, or origin. He wins the fight on her behalf, and, restored to monarchical legitimacy, she accepts his hand in marriage. But despite the overwhelming and immediate evidence that the hero is worthy of trust, the seeds of distrust are quickly planted in her mind. Consumed by doubt, Elsa finally demands to know the stranger's identity. He tells her that he is Lohengrin, son of the Grail King Parsifal, sent from heaven to rescue her. And since she broke her vow, he must irrevocably depart. Michael Tanner has argued that Elsa's behavior exemplifies a profound insight on Wagner's part into the nature of trust (M. Tanner, *Wagner* [Princeton University Press, 2002], chapter 6).

This behavior is genetically programmed in social animals, including humans. At first blush, this seems to be an entirely unrelated basis of trust. But evolutionary biologists explain Empathy (along with altruism and other collaborative behaviors) in exactly the same language that game theory uses for reciprocity. Individual organisms that "trust" each other through the exchange of these signals will gain when interacting with each other but lose to opportunistic organisms that take advantage of that "trust."[9] However, since both trust and opportunism are genetically programmed behaviors, competition is actually occurring across gene pools, not between organisms. And in some contexts, such as herds, the collaborative gene will win out.[10] Empathy is thus the emergent collaborative equilibrium of repeated prisoner's dilemma games ***among genes***, and thus a form of reciprocity. We humans, bearing those genes, are thus programmed to feel empathetic trust in a single encounter.

Shared Identity. Shared Identity is trust premised on the trustor and trustee, ***both*** being members of an exclusionary group variously defined not only by norms and values, but by affiliation, shared experience, sympathy, and common interests, purpose, or enemies. Like Empathy, the propensity to tribalism is biologically grounded, though the definition of the tribe is a contingent social (sometimes political) fact.[11] Nations, military organizations, and business teams all reinforce this sense of shared identity through rituals, rites of initiation and passage, anthems, symbols, uniforms, and myths.

Again, although these two bases are logically distinct, they are frequently and powerfully coincident: from corporate teams sharing a late-night pizza to army boot camp, a sense of shared identity is deeply grounded in close personal interactions.

Relationship. Relationship is a combination of friendship, intimate mutual knowledge, a history of mutual favors, shared experiences, and common endeavors. It is thus based both on reciprocity (the value each party rationally sees in ***future*** collaborative transactions) and on mutual perceptions of trustworthiness (attribution of values and goals to the counterparty based on ***past*** experience). Relationships (including, at the apex, loving relationships) are the most powerful and enduring basis of personal trust, but also the slowest and most expensive to establish.

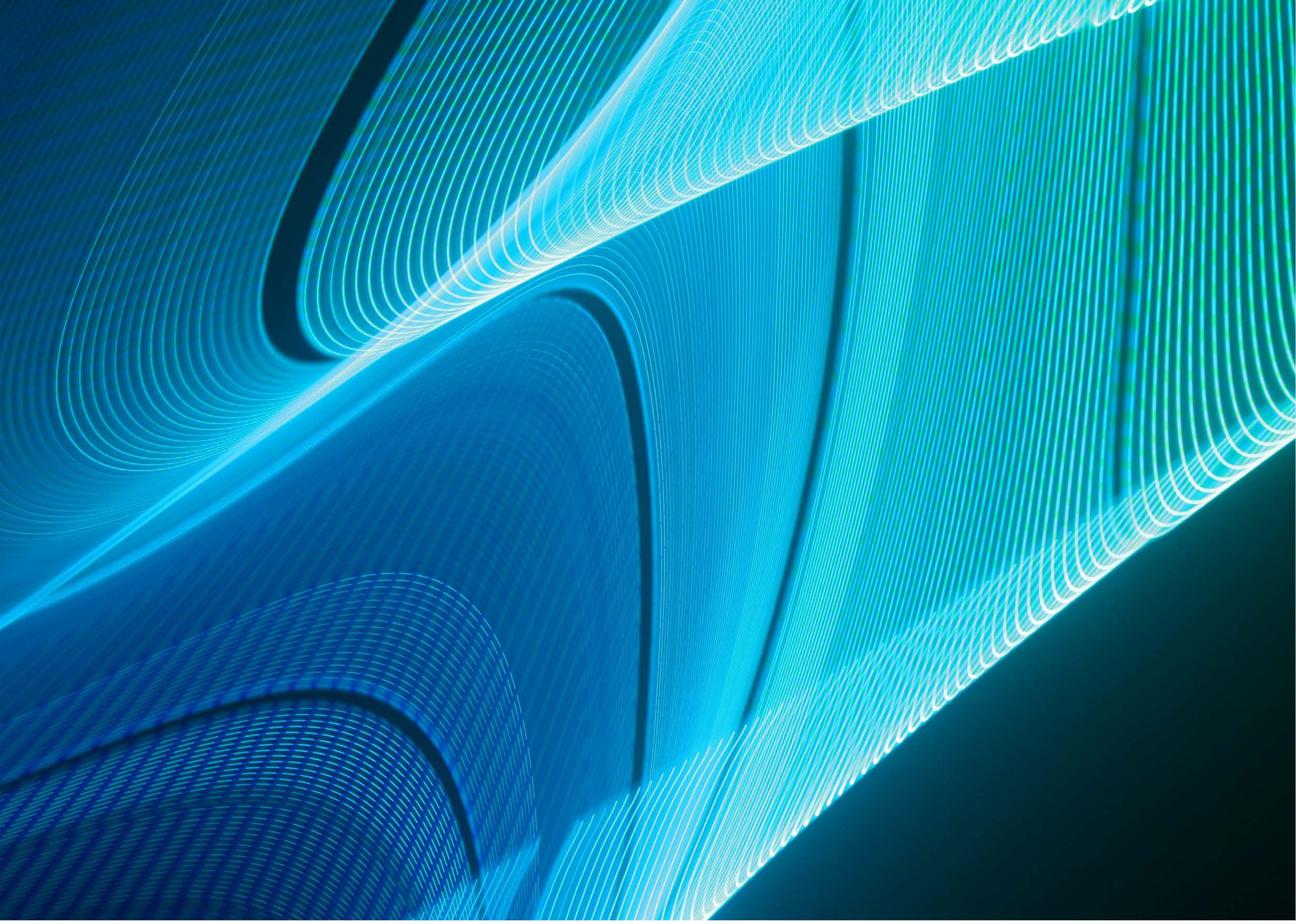
Relationships are also crucial among corporations. Unlike General Motors or Ford, which traditionally relied on the purchasing department to administer arm's-length competitive bidding processes, Toyota has long emphasized close and committed collaboration with its suppliers. This extends deep into factory-floor collaboration and continuous sharing of production improvements. In 1997, a fire in the factory of Aisin Seki, a component supplier, threatened to shut down the entire Japanese Toyota production system for months. Dozens of companies collaborated, improvised, and "swarmed" over the problem, trusting and trusted to act without instructions, contracts, or compensation. The first replacement parts were delivered in 85 hours; in two weeks, the assembly lines were back to full production. The trust-centric model of the Toyota Production System has become a model for supply chain management.[12]

[9] This slightly specialist use of the word "opportunistic" is owing to Oliver Williamson (*Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Internal Organizations* [Free Press, 1975]). In the context of incomplete contracts, he defined opportunism as "self-interest-seeking with guile."

[10] This view of selection occurring at the gene level rather than the organism was popularized by Richard Dawkins in *The Selfish Gene* (Oxford University Press, 1976). The remarkable application of game theory to evolutionary biology is put forward by John Maynard Smith. See J. M. Smith and G. R. Price, "The Logic of Animal Conflict," *Nature* 246, no. 5427 (1973): 15–18, and J. M. Smith, *Evolution and the Theory of Games* (Cambridge University Press, 1982). See also R. Trivers, "The Evolution of Reciprocal Altruism," *The Quarterly Review of Biology* 46, no. 1 (March 1971): 35–57.

[11] See C. Clark, B. Liu, and B. Winegard, "Tribalism Is Human Nature," *Current Directions in Psychological Science* 28, no. 6 (2019), https://journals.sagepub.com/doi/10.1177/0963721419862289.

[12] The story of the Aisin fire is narrated in P. Evans and B. Wolf, "Collaboration Rules," *Harvard Business Review*, July–August 2005.



How Trust Is Shifting Organizational Boundaries

Historically, in the “analog” world, humans have developed a variety of mechanisms to address the trust problem: mechanisms such as honor codes and face-to-face meetings (to generate trust) and civil law and double-entry bookkeeping (to manage distrust). But in the mechanisms that generate trust, we confront a fundamental tradeoff between strength and scalability: the strongest bases of trust develop only in small, typically in-person, groups. This tradeoff results in three types of analog trust in the modern market economy. Small transactions premised mainly on Relationship and Empathy occur within co-located groups. Large transactions occur among people and organizations managing mutual distrust through contracts and the law. And individuals trust large corporations through broadcast communication, primarily brand advertising (a form of Reputation-at-Risk). Conspicuous by its absence is the trust that enables small transactions among counterparties that lack a prior relationship.

Trust within an organization is mediated in much the same tripartite manner: small-scale mutual relationships among colleagues who work together; authority imposed through the corporate hierarchy to coordinate collaboration on a larger scale (with the ultimate sanctions of promotion or dismissal); and perhaps efforts to forge Shared Identity through broadcast slogans, pep rallies, uniforms, and videos on company purpose—all designed to make employees feel they are part of something bigger and nobler.

Nobel prizewinning economist Ronald Coase first proposed that the boundary between markets and organizations is determined by transaction costs. Hierarchical organizations, he argued, transact more cheaply than do markets. But as organizations grow larger, they become less flexible. So, the boundary of the corporation is set at the point where the marginal cost of inflexibility exceeds the marginal benefit from cheaper transactions. Transaction costs include the cost of search, negotiation, payment, and settlement. But most of these have been driven toward zero by the frictionless information economics of the digital economy. There is less and less for organizations to economize on and therefore less and less need for a hierarchical organization. Markets have substituted for traditional hierarchies. But *trust, too, is a transaction cost*: there is a cost to establishing it (say, in building a relationship) and a cost to managing around its deficiency (say, in negotiating a contract). And it is cheaper to create trust and manage distrust inside the corporation than it is with outside suppliers and customers. This particular transaction cost has thus *not* been reduced to vanishing by technology. Indeed, like a rock exposed by the ebbing tide, trust emerges as the dominant, residual transaction cost as those other costs recede. And therefore—*pace* Coase—the economics of trust are increasingly the principal determinant of the corporation's boundaries.

Indeed, one of the most important developments in modern capitalism has been the development of the ecosystem: "a dynamic group of largely independent economic players that creates products or services that together constitute a coherent solution."[13] Ecosystems fall midway between classical hierarchies and classical markets. What sets ecosystems apart are the mechanisms by which rich collaboration is achieved, of which there are two. One is the (generally trusting) relationships that enable resource sharing, such as those within the Silicon Valley community and the Toyota Production System. The other is the structuring of interfaces by a central orchestrator, as in the Apple iOS or Windows app developer communities. A fundamental reason for ecosystems' development is that they enable new solutions to the problems of managing trust.[14]

There is much at stake here:

> **The economics of trust are key to the success and scalability of markets, organizations, and ecosystems. And the economics of trust increasingly determine the boundaries between these alternative modes of economic organization.**

So, how does technology shift these economics?

[13] U. Pidun, M. Reeves, N. Knust, "How Do You Manage a Business Ecosystem?" BCG Henderson Institute, January 2021, https://www.bcg.com/publications/2021/how-to-manage-business-ecosystem.

[14] See also P. S. Adler, "Market, Hierarchy, and Trust: The Knowledge Economy and the Future of Capitalism," *Organization Science* 12, no. 2 (March–April 2001): 215–234.

The Digital Trust Network as a Stack

Complex technologies are organized in a “stack”—a modular architecture in which general-purpose, stable, lower-level functions provide enabling services to specialized, adaptive, higher-level functions.[15] The whole system is coordinated through interoperable interfaces. In this technical sense, the internet, for example, can be considered a stack.

But so is the service-oriented architecture of Amazon.com: cloud data systems support its warehouse fulfillment systems and merchandising systems. Those merchandising systems, in turn, support apps on devices. Internally, each of these functions is a set of human organizations. (Indeed, vendors using Amazon Fulfillment or corporate customers of Amazon Web Services are “modules” in the Amazon stack but outside the company.) So, a stack in an industrial context is a set of functional modules that interact in a hierarchical fashion using standard digital interfaces. Internally, those modules may be just software (such as a blockchain), but often they are complex human organizations—divisions of a corporation or independent enterprises coexisting in a common ecosystem.

[15] A stack does not require that *all* higher-level functions depend on *every* lower-level function but does require that *no* lower-level function depends on *any* higher-level function. This is exemplified in the discussion that follows. The evolution and logic of stack architectures in software is brilliantly described by Carliss Baldwin and Kim Clark in their monumental *Design Rules, Vol. 1: The Power of Modularity* (MIT Press, 2000).

A digital trust network is a stacked system architecture that enables the five trust-generation bases discussed earlier along with the complementary management of *dis*trust. It is embedded in the more complex architecture that defines an entire market, enterprise, or ecosystem. Abstracted from that larger system, it has seven modular mechanisms: Media, Authentication, Data Control, Social Networks, Aggregation, Constraint, and Authority.[16] (See Exhibit 2.)

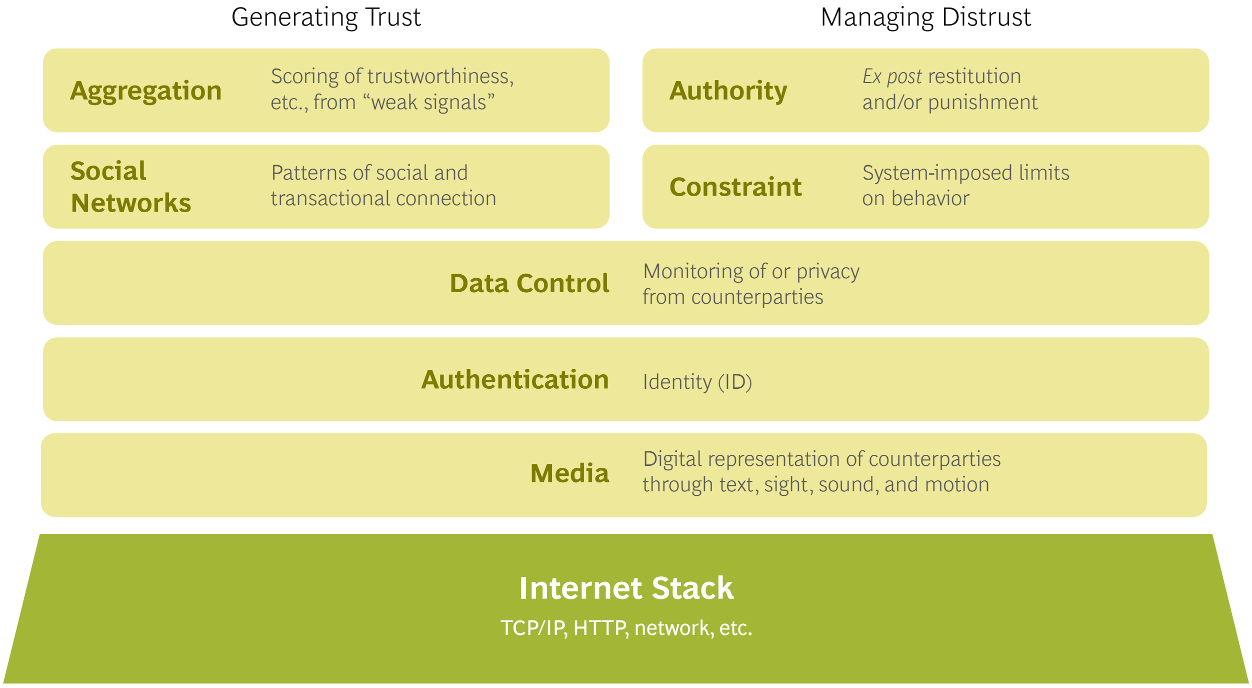
A DTN is a stack in the specific sense that lower-level mechanisms provide services to enable higher-level mechanisms, but not the reverse. At the bottom of the stack, functionality is quite general purpose (and indeed supports capabilities unrelated to trust, such as TCP/IP). At the top of the stack, functionality is bespoke: eBay uses one reputation mechanism, Etsy another.

Three of these mechanisms—Media, Authentication, and Data Control—may engender trust in their own right, but they are also critical enablers for other mechanisms that generate trust and manage distrust. As we'll see, two of these mechanisms—Social Networks and Aggregation—enable counterparties to trust each other. Two others—Constraint and Authority—allow parties to manage around their lack of trust.

Now, many of the providers or operators of these modular mechanisms are agents in their own right that play a role in an individual transaction as "intermediaries." Intermediaries, in general, need to be trusted. Trust of an intermediary is built from the same seven mechanisms, possibly requiring the trust of further intermediaries. For example, guests trust the Airbnb host (say, for the cleanliness of their rental) because of the intermediation of Airbnb; they trust Airbnb (say, for the integrity of payment) because of their bank's intermediation; they trust the bank (for the security of their funds) because of the Federal Reserve's intermediation, and so on. In this sense, the architecture of trust is often nested.

These seven mechanisms are variously affected by digital technologies: some tangentially, some fundamentally. Technology lowers transaction costs and increases the scalability of trust mechanisms, enabling trust to be generated or managed for smaller, intermittent transactions among parties with weaker ties. It drives the substitution of trust for distrust and drives the substitution of one basis of trust for another. The workings of markets, organizations, and ecosystems are changed, as are the economic boundaries between them.

Exhibit 2 - The Basic Architecture of a Digital Trust Network


Generating Trust

Managing Distrust

Aggregation Scoring of trustworthiness, etc., from "weak signals"

Social Networks Patterns of social and transactional connection

Authority *Ex post* restitution and/or punishment

Constraint System-imposed limits on behavior

Data Control Monitoring of or privacy from counterparties

Authentication Identity (ID)

Media Digital representation of counterparties through text, sight, sound, and motion

Internet Stack TCP/IP, HTTP, network, etc.

Source: BCG Henderson Institute analysis.

[16] Throughout this discussion, we use the very general word "mechanism" to describe the modular functions of the DTN and the systems, organizations, or communities that variously perform them. Mechanisms are observable processes, distinct from the bases of trust discussed earlier; bases are thought processes in the mind of the trustor.

Media

Media—email, text messages, encrypted PDFs, JavaScript in a web page, the HTTPS protocol, voice, cameras—enable all other trust mechanisms. Since the Bronze Age, merchants worldwide have imposed Constraint on each other by notching and then splitting a tally stick as an irrevocable record of their transaction; in medieval England, parish churches kept an accessible register of births, marriages, and deaths to certify Identity.

> **In our digital era, almost every mode of trust creation or distrust management depends at bottom on media. Face-to-face eye contact and the handshake are really the only exceptions.**

Media not only enable other trust mechanisms, but they can also enable trust directly through Empathy. Media's ability to generate Empathy is not new: when Rudolph Valentino died in 1926, a despairing crowd of 100,000 fans rioted at his funeral home in New York, and some half-dozen heartbroken women committed suicide.[17] More prosaically, candid writing, a casual snapshot, or a whimsical 15-second video can facilitate the nuanced communication of "honest signals" across distance, building trust even among people who have never met in person.

The Zoom Boom and Empathy. The experience of COVID lockdown has precipitated a step change in our willingness to rely on digitally intermediated Empathy: the "Zoom Boom." To a surprising extent, people have learned (perforce) to trust each other remotely through video streaming. Platforms have launched clever features to bring users closer together: the Duets feature on TikTok, for instance, assembles separate feeds to display people lip-synching to the same music. Over the longer term, tech companies and telcos are investing billions in the provision of lower-latency, higher-resolution video, and—with spatial audio, virtual reality, and augmented reality—evolving from two dimensions to three. The I/O device is becoming progressively more intimate: from desktop, to phone, to goggles, to "smart glasses"—and by the end of this decade, some predict, augmented-reality contact lenses. The more intimate the device, the greater the extent to which it is portable and ever-present. Microsoft, in its labs, has demonstrated a technology the company calls "Holoportation," in which people can interact remotely by encountering, through HoloLens goggles, convincing avatars of others presenting in their own physical space. And there is much experimental evidence that perceiving someone whole-body, in the round, in a shared physical context, enhances Empathic trust.[18] If that is experimentally feasible today, it will be cheap and irresistible in a decade.[19]

Media don't simply entail people trusting others through the digital relay of honest signals; people can also trust—and be trusted by—machines.[20] The fusion of AR, natural language processing, and the animatronic technology used in Hollywood enables machines to present a "human face" quite literally: a digitized avatar. The Royal Bank of Scotland already employs a winsome "trainee" called Cora to answer customers' queries in its NatWest branches in London. Cora (developed by Soul Machines in collaboration with IBM) engages not only in moderately intelligent conversations (the "Turing test") but appears head-and-shoulders on a screen, makes eye contact, and communicates some of the nuances of Empathy.[21] As with Holoportation, what matters is not the limited capabilities of such technologies today, but the path of exponential improvement on which they are so obviously launched.

AI as a Trustworthiness Detector. In the reverse direction, from machine to human, neural networks can recognize the subtleties of human emotion through language, head orientation, gesture, vocal timbre, and attention. AI can construct a real-time mesh model of a subject's face and associate the resulting patterns with half-a-dozen universal emotions, achieving correlations as high as .88 with self-reported emotional states.[22] These techniques are used mainly by marketers, usability engineers, and educators who are trying to improve the subject's experience with a product or service, and where the laboratory subjects know they are being observed and issues of trust do not arise. But it is only a small leap to apply these techniques in contexts where trust is at issue. AI can sometimes predict whether a person is trust*ing*; for example, in laboratory experiments, individuals experiencing disgust (as measured by AI) judged other people in the study as less trustworthy and were less likely to lend them money.[23] AI can also be used to predict whether a person should be trust*ed*—by attempting to distinguish, for example, true from faked emotions.[24] So we are reaching the point where sentiment analysis and lie detection can be embedded in every digital interaction, with or without the subject's

[17] J. Killerlane, "These Women Were So distraught After Rudolf Valentino's Death That They Committed Suicide," February 3, 2018, https://historycollection.com/women-distraught-rudolph-valentinos-death-committed-suicide/2/.

[18] See, for example, *Stanford News*, October 17, 2018, https://news.stanford.edu/2018/10/17/virtual-reality-can-help-make-people-empathetic/.

[19] For detail on Microsoft's project, see https://www.microsoft.com/en-us/research/project/holoportation-3/. AR is widely perceived among the tech giants as the next general digital interface and therefore a make-or-break priority. According to *The Verge*, Apple senior executives believe that AR "could supplant the iPhone in a decade"; see https://www.theverge.com/2019/11/11/20959066/apple-augmented-reality-ar-headset-glasses-rumors-reported-release-date. Facebook has similarly prioritized Oculus, and Google is a major investor in Magic Leap.

[20] More precisely, since machines are not "parties," we should refer to the programmer or owner of the machine as the trustor or trustee.

[21] See https://www.soulmachines.com/2018/02/press-natwest-begins-testing-ai-driven-digital-human-in-banking-first/.

[22] S. Stöckli et al., "Facial Expression Analysis with AFFDEX and FACET: A Validation Study," *Behavior Research Methods* 50, no. 4 (2018): 1446-60.

[23] T. Kugler et al., "On Trust and Disgust: Evidence from Face Reading and Virtual Reality," *Social Psychological and Personality Science* 11, no. 3 (July 1, 2019), https://doi.org/10.1177/1948550619856302.

consent. The ethical issues in this type of application are obvious and have provoked considerable debate among researchers in facial recognition.[25]

Authentication

The internet protocol is stateless: there is no built-in retention of information from one data request to the next. Digital identity (in the mundane sense of "ID" or who you are) is not an automatic correlate of bodily extension in space and time as it is in the physical world.[26] And without continuity, there is no "shadow of the past" (trustworthiness) or "shadow of the future" (reciprocity).

> **The entire edifice of digital trust therefore critically depends on some overlay for establishing ID.**

ID is nested: most forms of ID depend on a secure binding to another ID. At the root is something like a birth certificate or passport, generated in the physical world. Take India's Aadhaar program, by far the most ambitious digital identity program in the world. (Aadhaar means "foundation" in Hindi.) The program requires that each citizen appear in person at an enrollment center with a sheaf of original paper documents proving date of birth, address, family relationships, and so forth. Biometrics are measured on the spot: fingerprints, signature, iris scan, and photograph. The digitized information is recorded in a database which can then provide authentication from combinations of a physical card, password, and biometric markers. On that "identity layer," the government envisages an entire "India stack" of trust-related services.[27] ID also applies to things: consider DeBeers's Tracr platform for diamond tracking, in which the diamond's "identity" is established at the very beginning of the value chain to ensure authenticity.

Building on this physical root is an increasingly elaborate hierarchy of partially interoperable digital IDs. Because these are prime targets for identity thieves, the security bindings between physical and digital (and between one digital ID and another) have become an extraordinarily elaborate thicket: longer and longer passwords, heavy encryption, multifactor authentication, biometrics (fingerprint, face, voice, retina), VPNs.

Logically, a single, absolutely secure, unique, digital ID should suffice for all purposes.[28] That would be the rational approach if one were starting from scratch (as with Aadhaar). In practice, there is a hierarchy of digital IDs issued by multiple, interdependent, and overlapping entities (banks, social security administrations, hospitals, municipal registries, licensing authorities), at some cost to efficiency and consistency. But in reality, this diffusion of trust authority minimizes the chances that some rogue government official will erase one's entire identity (as happened to the hapless Will Smith in the 1998 movie *An Enemy of the State*). And although individual IDs can be compromised, the multiplicity of IDs gives the overall system a measure of robustness against both catastrophic failure and deliberate sabotage. Proponents of blockchains argue that a single crypto architecture would place control of identity uniquely and irrevocably in the proper hands: those of the subject. But as with so many other dimensions of trust, in advanced economies, the current institutional arrangements, while duplicative and imperfectly interoperable, are generally good enough to preclude disruption.[29] In some emerging economies, it's a different story.

Data Control

Data Control—by which we mean the control ***of*** data, not control of the counterparty ***through*** data (which is a constraint, discussed later)—has two aspects: the ability to monitor (the trustor's ability to get information about the trustee) and the ability to prevent monitoring (the trustee's ability to prevent the trustor from obtaining such information). Data Control is deeply political—it's about who has power over the flow of information. We may feel one way about governments or corporations monitoring citizens, another about citizens monitoring governments or corporations, and perhaps differently again when the monitoring is symmetrical (as when the US and Soviet Union used spy satellites to verify each other's missile deployments).[30] When we approve of the monitoring, we speak of "data transparency"; when we disapprove, we speak of "a violation of privacy rights."

[24] H. Ugail and A. Al-dahoud, "A Genuine Smile Is Indeed in the Eyes—The Computer Aided Non-invasive Analysis of the Exact Weight Distribution of Human Smiles Across the Face," *Advanced Engineering Informatics* 42, October 2019.

[25] See, for example, R. Van Noorden, "The Ethical Questions That Haunt Facial-Recognition Research," *Nature* News Feature, November 18, 2020, https://www.nature.com/articles/d41586-020-03187-3.

[26] As we use the term, "ID" is the contents of one's birth certificate or passport. It is not to be confused with "Identity," meaning membership of a social category or a socially distinguishing feature in which a person takes special pride. Shared Identity in this latter, cognitive sense is one of the five bases generating trust.

[27] This is richly described in "What Is IndiaStack?", https://www.indiastack.org/about/.

[28] Since there are multiple governments, maybe in an interoperable format comparable to the International Civil Aviation Organization's standardization of physical, machine-readable passports implemented in the 1980s.

[29] For example, the Known Traveler Digital Identity system proposed by the World Economic Forum; see http://www3.weforum.org/docs/WEF_KTDI_Specifications_Guidance_2020.pdf.

Monitoring and Transparency. The relation between Data Control and trust is quite intricate. Monitoring reduces asymmetries of information and thus uncertainty about the counterparty's behavior. The first-order consequence is that the trustor has therefore less need to trust. If monitoring indicates opportunistic behavior by the trustee, the trustor can terminate the relationship (abort future transactions), blacken the trustee's Reputation-at-Risk, or appeal to Authority for restitution—for example, in a court of law.[31] Indeed most forms of Authority (discussed below) depend intimately on means of monitoring. Knowing all that, there is a second-order effect: the trustee subjected to monitoring may be motivated to behave in a more trustworthy manner. Knowing and seeing that, the trustor has better reason to trust. This was Ronald Reagan's famous but slightly paradoxical slogan, "Trust but verify."[32] Jeremy Bentham's 1791 concept of the "Panopticon" prison was predicated on the idea that if prisoners thought they *might* be under surveillance but had no way of knowing whether indeed they were, they could be trusted.[33]

Governments see surveillance as a key tool for controlling crime and other undesirable behaviors. The number of surveillance cameras worldwide will exceed 1 billion by the end of 2021; over 50% of them are in China's Skynet system.[34] Those selfsame AI systems that can read human Empathy can identify the license plates of scofflaws at expressway tollbooths or the faces of jaywalkers on the streets of Shanghai.

In high-risk, low-trust situations, employers have long aggressively monitored their employees: strip searches of workers as they finish their day's labor in African diamond mines, or the "eye in the sky" installed behind two-way mirrors in Las Vegas casinos.[35] But technology has enabled a massive escalation in the universality and intrusiveness of employee monitoring. Based on continuous identification via logins, facial recognition, and surveillance cameras, corporations can track movements, distractions, and productivity by analyzing keystrokes and the duration of phone calls and bathroom breaks. During the COVID epidemic, a number of colleges required online students to install software that monitors test-taking violations such as referring to textbooks, notes, or Wikipedia.[36] Especially in customer service and telemarketing, human monitors (invisible and randomized, as Bentham would have applauded) eavesdrop to assess employee time-wasting, friendliness toward customers, and conformity to standardized scripts. Increasingly, this monitoring can be performed by machines.

Monitoring can benefit all parties in profound ways. Google and Facebook assert correctly that they are able to place more relevant advertising on the basis of the tracking technology they employ to infer the user's intent and interests. The problem is that even if that is broadly true, many users do not trust how big tech companies handle their personally identifiable information. So, rightly or wrongly, monitoring in one direction provokes distrust in the reverse. Companies such as Apple, whose business models do *not* depend on such monitoring, have made their "respect for privacy" a hallmark of why their brand especially merits trust.

Preserving Privacy.

> **A major frontier in trust technology is therefore to finesse this tradeoff: to enable useful insights based on the power of AI applied to aggregated data while preserving the privacy of individual data subjects.**

An extreme case is medical research, where deep insights are to be harvested from comprehensive statistical analyses of symptoms, genomes, treatments, and outcomes, but the centralization of patients' medical information in these so-called "disease registries" is highly problematic.[37] In some contexts, anonymization suffices, but it is remarkably easy to reconstitute identity from individually anonymized data points.[38] An approach finding increasing favor among medical institutions is "federated learning": a machine-learning technique in which an algorithm is trained on data from multiple data sets without those data sets ever being merged; instead of the data flowing to (the

[30] For a searing exposition of this issue, see S. Zuboff: *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (PublicAffairs, 2019).

[31] The ability of Data Control to manage distrust is thus contingent on the ability of the trustor to link it to another of the seven mechanisms.

[32] Ironically, the phrase is a *Russian* proverb: "Doveryay, no proveryay."

[33] One Panopticon was actually built: the Presidio Modelo in Havana. Fidel Castro was incarcerated there. It is now a museum.

[34] This projection by IHS Markit was reported in the *Wall Street Journal*, https://www.wsj.com/articles/a-billion-surveillance-cameras-forecast-to-be-watching-within-two-years-11575565402.

[35] The role of monitoring in securing trust in casinos is memorably described by the voice-over by Sam "Ace" Rothstein (played by Robert De Niro) in Martin Scorsese's 1995 movie *Casino*: "In Vegas, everybody's gotta watch everybody else. Since the players are lookin' to beat the casino, the dealers are watchin' the players. The boxmen are watchin' the dealers. The floormen are watchin' the boxmen. The pit bosses are watchin' the floormen. The shift bosses are watchin' the floor bosses. The casino manager is watchin' the shift bosses. I'm watchin' the casino manager. And the eye in the sky [surveillance camera] is watchin' us all." (https://www.youtube.com/watch?v=aIPmu6bYZOs)

[36] Proctortrack (https://www.proctortrack.com), for example, claims to identify up to 17 distinct test violations, including impersonation, unauthorized use of a browser or other apps, leaving the room during the test, presence of others in the test-taking room, and so forth. The accuracy and equity of these technologies has been roundly criticized.

[37] Predictably, the greatest progress in this use of "registries" to aggregate patient data for the purposes of medical research has been made in comparatively high-trust societies such as the Scandinavian countries; see https://www.bcg.com/publications/2011/health-care-payers-providers-public-sector-value-based-health-care-interactive.

researcher estimating) the algorithm, the algorithm is distributed among (the institutions guarding) the data. In a similar fashion, Google has proposed a technology called FLoC (Federated Learning of Cohorts) by which individual browsing behavior is tracked locally, and only aggregated data is uploaded to Google for the purposes of placing advertisements. As users have become distrustful of (and platform providers such as Apple overtly hostile to) the cookies and apps that track user behavior, solutions such as FLoC are key to preserving the users' trust in advertising-based business models.

The technologies that enable surveillance also enable "sousveillance," or observation from below.[39] There is a widespread perception in the US that Minneapolis police officer Derek Chauvin would not have been convicted of the second-degree murder of George Floyd without video recorded on mobile phones by bystanders. In an institutional context, trusted intermediaries such as accountants perform sousveillance of corporations on behalf of investors. Human Rights Watch performs sousveillance of prisons and refugee camps across the world. As investors and consumers become increasingly insistent that corporations fulfill environmental, social, and governance (ESG) goals such as carbon neutrality or diversity, their need for trusted third-party measurement and validation will become increasingly urgent.

An especially critical issue in supply chains is consumers' growing concerns about the provenance of and labor conditions under which the ingredients or components of items they buy are produced. In complex international supply chains, that issue is totally opaque. But if there is a trusted intermediary certifying the "fair trade" status of, say, a consignment of coffee, it may be feasible to construct an end-to-end DTN that delivers that trusted reassurance to the consumer at the point of purchase. This is sometimes cited as an application for blockchains, and solutions are being developed along those lines. But the hard problem is establishing trust of, and a business model for, the certifier of the original production conditions; given that, the rest is data logistics. Blockchains are, in fact, limited. Every time a blockchain interfaces with the physical world—whether in the creation of a diamond's digital twin (in the case of DeBeers) or the placement of a tracking sensor in a food shipping container—it requires a human "oracle" or some type of IoT, managed by a human operator or organization, to capture that information.

These frontiers all illustrate that Data Control—both monitoring and privacy—is a domain of trust management where digital technology is radically transformative. Businesses will be built, and competitive advantage established, by addressing new needs and massively scaling traditional solutions. But political controversy, implicit in so many of our examples, is never far from the surface.

Social Networks

Obviously, social networks are as old as humanity itself: they are the template of relationships, of course, and the locus of social capital—a pervasive, if ill-defined, basis for trust. Digital social networks are patterns of human connection intermediated by digital media. Platforms such as Facebook and Twitter have a business model centered on curating such networks, but curation is not necessary: social networks can also be mediated by email, texting, or Zoom. Curated social networks require Authentication, generally by name (although pseudonyms are sometimes sufficient).

In the physical world, connections are made and reinforced through in-person encounters, each requiring proximity and some investment of time. In the world of digitally intermediated social networks, encounters are unbounded geographically and can be as brief as a "like," emoji, or tweet. So, while the affective content of the encounters may be attenuated (but improved by the advances in digital Empathy discussed earlier), their span and frequency can be enormously increased. Technology might therefore be expected to enlarge individuals' social networks and therefore expand Relationship trust.

However, this appears not to be the case. The most comprehensive analysis of network patterns for people who are digital social networkers is by Oxford's Robin Dunbar.[40] He uses an established classification of three degrees of closeness: "support clique" (the people on whom one depends for emotional support in a crisis); "sympathy group" (close friends); and "friends" (equated with "friending" on Facebook).[41] These categories reflect interaction frequencies as well as emotional intimacy. Before the advent of digital social media, studies indicated that the average numbers of people in each of the three categories in a person's "egocentric network" were 3.8, 11.3, and 150, respectively (with high variance).[42] In Dunbar's large, stratified samples (one of "frequent social networkers" and one of "business

[38] In 1997 Latanya Sweeney, then a graduate student at MIT, matched Massachusetts Governor Bill Weld to his own medical record, reconstituting the link entirely from anonymized public sources. She went on to propose the concept of *k*-anonymity: a statistical measure of the degree of privacy afforded by a data set. Since then, a large mathematical literature has evolved that attempts to quantify and optimize the tradeoff between the statistical acuity of a data set and its degree of anonymity. The deliberate introduction of "noise," for example, lowers acuity but increases privacy in predictable ways.

[39] The felicitous term "sousveillance" was coined by Steve Mann, professor in the Department of Electrical and Computer Engineering at the University of Toronto. He has been wearing computerized eyewear for some 35 years.

[40] R. I. M. Dunbar, "Do Online Social Media Cut Through the Constraints That Limit the Size of Offline Social Networks?" *Royal Society Open Science* 3, no.1 (January 2016): 150292, https://royalsocietypublishing.org/doi/10.1098/rsos.150292.

[41] Surveys have employed two further categories: acquaintances ("people we know well enough to have a conversation with") and "people whose faces we can put a name to" (averaging 500 and 1,500, respectively). These were not the subject of Dunbar's 2016 research and are probably unrelated to relational trust.

professionals"), the numbers (for the first sample) are 4.1, 13.6, and 155. There's no significant difference. In both domains, there is a lot of variation (male versus female, introvert versus extrovert, young versus old, and so on) but no significant difference in the ***degree*** of variation. Dunbar concludes that connection at these more intimate trust levels is constrained by cognitive factors, not the physical ease of communication, so technology makes no real difference.

Social networks do little to strengthen or expand strong ties—those that underwrite relational trust. But they have had a big impact on the scale and power of weak ties, in particular those that influence Identity and Reputation.[43] Digital social networks facilitate the relay of short viral messages, or "memes." Platform operators try to maximize engagement by filtering and precisely presenting those memes that each member is most likely to "like" (and repost). This has the further merit of relieving the originator of any guilt about bothering their friends; the platforms do it for them.[44] Thus, memes spread among people who may barely know each other, binding them into a "community of attention." The corollary is "homophily": the tendency of like-minded people to connect with each other in comparatively dense clusters and thus reinforce shared perceptions and biases. Such clusters can become very large and geographically unbounded.

In consequence, groups defined by clusters in digital social networks have become increasingly a locus of Identity. And these clusters are no longer constrained by geography. Minority groups that could not achieve critical mass from physical co-location (around the neighborhood, at church, in clubs) achieve it via virtual connection: hobbyists, the LGBTQ community, expatriates, linguistic minorities, MMORPG (massively multiplayer online role-playing game) enthusiasts, Linux programmers, conspiracy theorists, and so on.

These groups often display "power-law" patterns in their internal connectivity: a pseudo-hierarchical organization of connectivity that emerges as a natural consequence of people desiring to connect to the well-connected. In such systems, a small number of individuals who devote a lot of time to networking are central and just a few degrees of separation from an immense audience.[45] Legitimized by the Reputation they thus enjoy within their group, they gain trust and thus become "influencers." There is a downside: influencers are gaining reputational trust largely at the expense of traditional brands, authorities, and news sources. In the context of the COVID epidemic, a recent study of content shared or posted on Facebook and Twitter found that 65% of anti-vaccine messaging was attributable to just 12 people, none of whom has relevant medical qualifications.[46] Corporate advertising strategies are being refocused to exploit this new channel as marketers pay influencers to promote their products. Partly in consequence, measured trust in traditional institutions, such as mainstream media and political institutions, is declining.

Aggregation

DTNs enable much larger numbers of casual interactions, both social networking and remote transactions. These weak links generate weak signals about trustworthiness. Statistics is in part the science of making strong inferences from large samples of individually weak signals: the "law of large numbers." Within DTNs, weak signals can be captured, aggregated, and analyzed to generate a robust digital Reputation, building, of course, on digital Media, Authentication, Data Control, and the various forms of social and transactional networks.

Conceptually, this aggregated data might be straightforward information about the trustee, communicated privately to the trustor, who therefore attributes norms of trustworthiness to that individual or enterprise. In the analog world, this process can take the form of chance comments and personal references. But it is far more powerful to broadcast the data digitally, thereby putting the trustee's reputation at risk. Everyone knows that everyone knows about the trustee's record. And this is what platform operators invariably do.

eBay pioneered digital reputation among a community of total strangers with its five-star system for seller ratings and feedback scores, a system that allows rating along

[42] R. A. Hill and R. I. M. Dunbar, "Social Network Size in Humans," *Human Nature* 14 (2003): 53–72, doi:10.1007/s12110-003-1016-y. An "egocentric" network is the set of people ("alters") with a direct tie to the subject ("ego"). Dunbar earlier speculated that the maximum size of egocentric networks in animals correlates with the size of their neocortex. This led to the proposal that human egocentric networks cannot exceed 150 (the so-called "Dunbar number"). Later research has indicated that the number 150 is better understood as a mean rather than an upper limit.

[43] Mark Granovetter famously made the distinction between "strong" and "weak" ties in "The Strength of Weak Ties," *American Journal of Sociology* 78 (May 1973): 1360–80.

[44] This is in contrast to, say, the telephone call before caller ID. At that time, a caller imposed on the recipient the cost of answering the phone whether or not the recipient wanted to communicate. Knowing that, callers (at least some of them) would hesitate to place the call. Digital media place control of communication entirely in the hands of the recipient: knowing that, and with zero marginal cost to messaging, senders have no reason not to send messages to everyone and allow recipients to decide whether or not they are interested in the communication. Spam is, of course, the abuse of this capability, but it is largely attributable to the failure of a lower-level function: Authentication.

[45] Mathematical "power-law" patterns of connectivity have been observed in many social networks, including blogosphere citations and hyperlinks among web pages. Réka Albert and Albert-László Barabási ("Statistical Mechanics of Complex Networks," *Reviews of Modern Physics* 74, no. 47 [2002]) developed an agent-based model of "preferential attachment" that predicts this larger pattern.

[46] The study was based on a sample of anti-vaccine content that was shared or posted on Facebook and Twitter 812,000 times between February 1 and March 16, 2021; see "The Disinformation Dozen," Center for Countering Digital Hate, https://252f2edd-1c8b-49f5-9bb2-cb57bb47e4ba.filesusr.com/ugd/f4d9b9_b7cedc0553604720b7137f8663366ee5.pdf.

multiple dimensions of performance. Most parents would be leery of leaving their child in the hands of a total stranger, but UrbanSitter invested heavily in trust mechanisms to build a thriving networking business. Not only do parents rate the nannies, but individual parents are able to review ratings by other identified parents whom they know. The rating system is aggregating strong signals, not weak ones.

The Bigger the Network, the More Powerful the Trust

The peculiarity of digital Reputation-at-Risk is its scalability: the larger the network of possible future transaction counterparties, the more powerful the trust engendered. More ratings provide the trustor with a larger, harder-to-manipulate sample, hence better estimates of trustworthiness in a narrowly statistical sense. But more importantly, they provide a larger audience, so the value to the trustee of their reputation score is proportionately greater: a fact of which the trustor is aware. Digital reputation allows an individual's "brand" to appropriate the scale of the network within which that reputation is embedded; a well-reviewed innkeeper on Airbnb can compete on trustworthiness with a national hotel chain.

Just how far this logic extends is uncertain. Many digital platforms were launched with the expectation that engineering Reputation-at-Risk via software alone would allow the business to avoid the costs and responsibilities of exercising Authority. But some have found that the limitations of reputational measurement—vulnerable to the infinite cunning of humans who game the system—leave gaps that only human adjudication can resolve.

Through clever algorithms (increasingly, AI), digital reputations can be extended from semi-objective criteria that everyone would agree on (honesty, timeliness, etc.) to entirely subjective dimensions of taste (a weaker form of trust, with respect to taste rather than truthfulness). This is done by appeal to homophily; Spotify and Netflix, for example, use this to make trusted recommendations on music and movies based on the "reputation" of the content in the eyes of like-minded consumers.

Democratizing Reputation Building

For better or worse, this tendency to homophily is driving the democratization of reputation: people are relying on the reputation established among large populations of amateurs rather than elite groups of professionals. Thus, Zagat competes with the Michelin Guide, and Rotten Tomatoes ratings are followed as widely as the reviews of professional film critics. A 2015 study found that IMDb votes were a significantly more important factor for predicting box-office success than the consensus of movie critics.[47] This is another way that digital technology is supplanting the trust traditionally enjoyed by establishment institutions.

The scalability of one's reputation within a DTN also implies economies of scale for the DTN platform itself: a larger platform can deliver more robust signals of trustworthiness for more trustees and, therefore, will be preferred by trustors. This (among other factors) has driven concentration among DTN platforms. As Reputation networks consolidate, so individuals cannot "escape" their reputation by migrating to a different DTN. This presents a fundamental tradeoff between Reputational trust and privacy. China's Social Credit System is a powerful platform for digital reputation intended to include every citizen and cover a very wide range of behaviors: unavoidability is a feature.[48] Europe has moved in the opposite direction, enshrining in its General Data Protection Regulation the "right to be forgotten."[49]

Now let's turn to the two mechanisms that allow for managing distrust.

Constraint

The need to engender trust can be reduced or eliminated by directly limiting the trustee's ability to act opportunistically. Rules, procedures, and protocols can be imposed by the trustor or by an agent operating on their behalf (who may themselves need to be trusted). These procedures compel the subject to behave in a trustworthy manner by limiting their scope to act otherwise. Digital Constraint is a large and expanding aspect of how corporations protect their assets and systems from external attack and from internal misuse; it is also a principal technique by which ecosystem platform operators minimize the need for participants to trust each other.

Moats and Internal Watchdogs

Traditionally, corporations have protected themselves from outside attack through a single security moat behind which any actor might then access or act upon any resource: once penetrated, the system is vulnerable. The modern architecture of so-called Zero Trust shifts access control from the perimeter to policies imposed on each individual (embedded in their devices) with respect to each resource. Firewalls, gateways, automated policies, and enforcement points thus provide "defense in depth."

[47] M. Wasserman, X. Zeng, and L. Amaral, "Cross-Evaluation of Metrics to Estimate the Significance of Creative Works," *PNAS* 112 , no. 5 (January 2015): 1281–86, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4321294/.

[48] A combination of government and business surveillance that assigns citizens a "score" based on their social and economic behavior. A low score, reflecting bad behavior, can restrict a person's ability to engage in certain transactions, such as purchasing plane tickets, buying property, or taking out loans.

[49] A concept hotly debated in the past decade that pits data privacy rights against free speech; it is the subject of Viktor Mayer-Schönberger's book *Delete* (Princeton University Press, 2011). See also https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-189; and J. Rosen, "The Right to Be Forgotten," *Stanford Law Review online* 88 (February 2012), https://review.law.stanford.edu/wp-content/uploads/sites/3/2012/02/64-SLRO-88.pdf.

That same Zero Trust architecture constrains employee behavior internally, not merely with respect to malicious acts, but any kind of nonconformity. Systems precisely define access privileges to secure facilities or to sensitive data such as source code, customer records, and passwords; they restrict employees' use of unauthorized software and nonstandard hardware; they strictly compartmentalize knowledge; they restrict emendation, copying, or distribution of sensitive documents; they impose templates and automatically reject incomplete applications or reports; they overrule excessively generous discounts offered by a salesperson or lax credit terms offered by a lending officer; they censor text that does not correspond to the officially sanctioned legal boilerplate.

Safeguarding the Ecosystem

In ecosystems, one of the most common applications of Constraint is in making transactions irrevocable.[50] Traditional escrow by notaries and attorneys performs that function, of course (at high cost), but now a cryptocurrency payment or blockchain smart contract can do that without human intervention. So does Uber when making escrow between driver and passenger, as described earlier.

But a platform manager such as Uber can do a lot more to constrain trustees' behavior. Its system also limits wait time, sets the fee at the outset, prescribes the route, and protects the pseudonymity of both parties.[51] The aim is to make the "contract" between driver and passenger as complete as possible in order to minimize the scope for opportunistic behavior.

Apple relies primarily on Constraint to manage trust within the iOS ecosystem. Apps can be installed only via the App Store, and payment can be made only through Apple Pay. Apps must be preapproved, conforming to Apple's definitions of security, quality, and decency. Viruses and malware, if they slip through that screen, are disabled as soon as they are identified. Apps are "sandboxed" so that they can access files and system resources only with the explicit approval of the user. With iOS 14.5, Apple required apps tracking users' device usage and interactions with other apps to obtain the user's explicit permission.[52] There is some controversy over Apple's motives for imposing these limitations, but whether users trust Apple or not (again, a separate question), it is striking that only 12% of users worldwide (4% in the US), when given the choice, allow app tracking.[53] Clearly, users do *not* trust apps gathering their personal information outside the necessary context of a direct transaction, and Apple (for whatever motive) has therefore done them the favor of constraining such behavior.

Constraint embedded in code denies the trustee the option to act opportunistically and can be deployed and upgraded at near-zero marginal cost across arbitrarily large systems. Monitoring, in contrast, can identify opportunistic behavior only after the fact and depends on complementary mechanisms to provide deterrence or restitution: breaking off a relationship, blackening a Reputation-at-Risk, or appeal to Authority. But that is for a single transaction: if the object of trust management is a transaction system (as is the norm), then Monitoring and Constraint become complements. Payment processors such as Visa and American Express, for example, continuously monitor their billions of credit card transactions for patterns of fraud, nowadays applying neural network techniques. Once a pattern is identified, the authorization rules are updated (within minutes) to constrain all future transactions. This kind of exponential learning characterizes the management of digital trust in general; nothing in the pre-digital world can match it.

[50] A crude but exotic pre-digital instrument for carrying out irrevocable transactions was the Glienicke Bridge over the Havel River, a checkpoint connecting West Berlin to Potsdam. It was used by the US and USSR for swapping spies, such as the 1962 exchange of U-2 pilot Gary Powers for Rudolf Abel, a Soviet spy. The bridge itself had long been closed and deserted. The KGB and CIA would bring their respective prisoners up to the bridge. The two spies would walk across, passing each other at the midpoint, thus facilitating a simultaneous and irrevocable transaction, requiring zero trust. (John le Carré memorably narrates such an exchange in his 1979 novel *Smiley's People*, and the exchange of Abel for Powers culminates Stephen Spielberg's 2015 film *Bridge of Spies*.)

[51] Theoretically, an Ethereum smart contract could perform all of these functions. The problem is that it would still be necessary to trust all the "oracles" submitting relevant information, and *by design*, the system would be extraordinarily difficult to upgrade. The Ethereum blockchain might be deemed more trustworthy than Uber's database, but there is scant evidence that distrust of Uber's recordkeeping is a problem. (Condemnation of Uber's corporate conduct is a huge problem, but that is a different matter.) Blockchain-based ride-hailing services have been variously proposed, but none, to our knowledge, has been funded, much less proved commercially viable.

[52] "User Privacy and Data Use," https://developer.apple.com/app-store/user-privacy-and-data-use/.

[53] T. Hardwick, "Analytics Suggest 95% of Users Leave App Tracking Disabled in iOS 14.5," May 7, 2021, https://www.macrumors.com/2021/05/07/most-iphone-users-app-tracking-opt-out/.

Authority

“Authority” is any legitimate system of rewards or punishments imposed on the trustee to sanction trustworthy behavior. Within the corporation, it is exercised through hierarchy; and across markets, through law. If Monitoring and Constraint, powered by technology, are the new approaches to trust management, Authority is the old way of doing things.

Authority can be imposed by the trustor or by some third party. It is exercised through a combination of rules (policies, laws, and precedents that ensure a modicum of consistency and predictability) and processes that interpret those rules and impose adjudications on the parties. The rules matter not just for fairness, but because consistency (and the certainty of being caught) enables deterrence. The maturation of legal systems (from Hammurabi and Justinian to the Code Napoléon) and the maturation of corporate hierarchies (from one-person caprice to by-the-book Weberian bureaucracy) is predicated on the premise that consistent rewards and punishment ***ex post*** motivate appropriate behavior (including trustworthiness) ***ex ante***.[54]

Where the Authority is exercised by the trustor, they have strong incentives to automate the process: hence automatic delisting of delinquent customers, denial of credit, rejection of obscene content or abusive blog posts, and so forth. Apple has deactivated nearly a quarter of a billion user accounts for fraudulent or abusive activity, including fake reviews.[55] But where the authority is a third party called in to adjudicate a dispute between trustor and trustee, automated arbitration is impossible, and technology becomes a weapon in the hands of the disputants. Adversarial processes—such as litigants facing each other in a court of law—are inherently zero-sum. Technology escalates capabilities on both sides: as a result, the net is not faster adjudication, but rather more complex arguments.[56] That is why the productivity of law has been so little enhanced by technology.

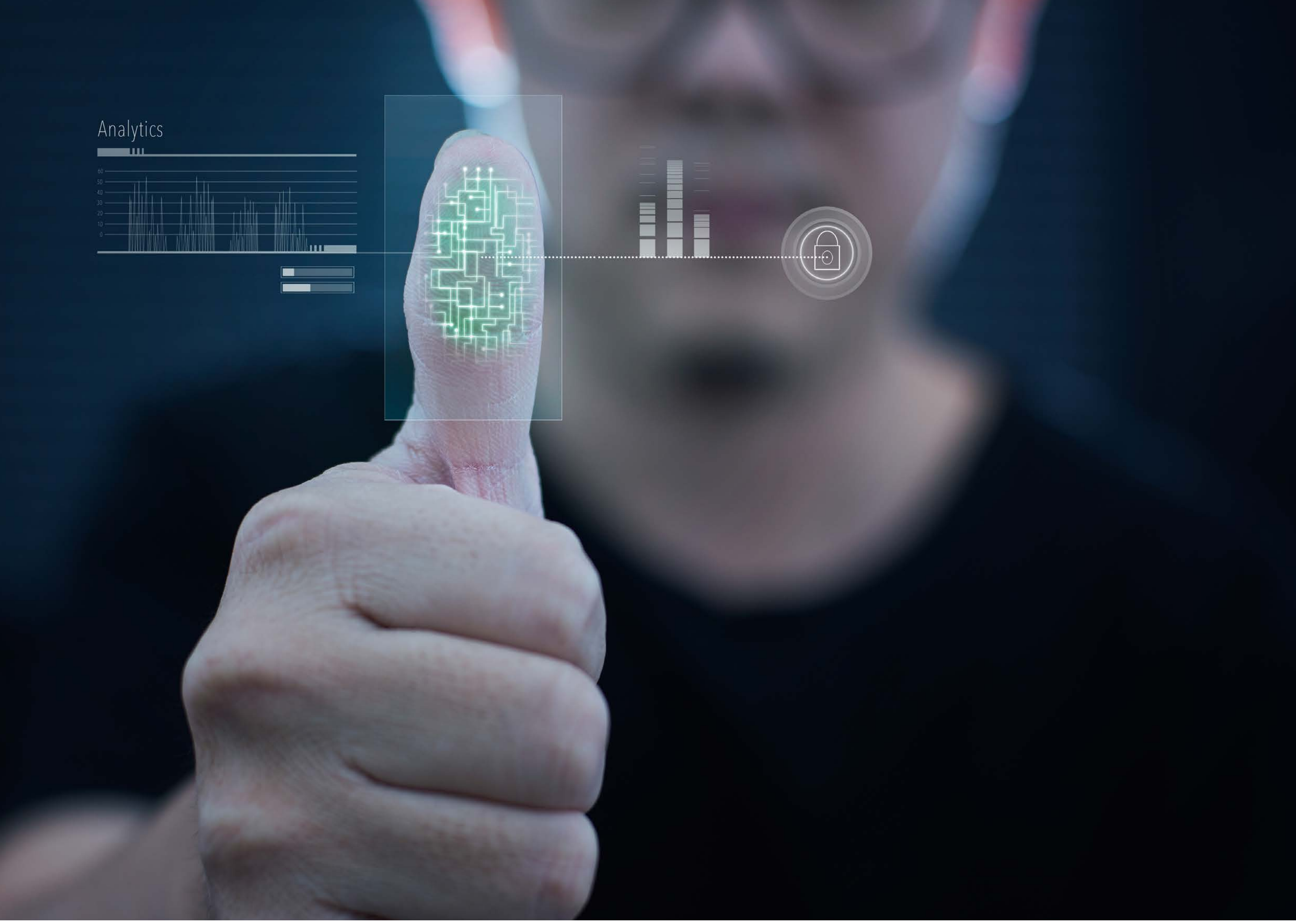
So, Monitoring and Constraint advance apace, as does Authority when it is imposed through digital mechanisms. But human adjudication languishes. In Larry Lessig’s celebrated trope, “software code” substitutes for legal code.[57]

[54] The use of Authority specifically for deterrence blurs on the edges the distinction between trust generation and management of distrust (the same is true with respect to monitoring). This ambiguity is implicit in Ronald Reagan’s use of the phrase “Trust, but verify”: does verification *enable* trust, or does it *eliminate the need* for trust? We categorize both Authority and Data Control as “managing distrust” because that is their first-order function: deterrence is second-order. Nothing in the larger argument hinges on this.

[55] They also rejected 424 million attempts to open accounts because of fraudulent and abusive patterns of behavior. These startling numbers are reported in Apple’s June 2021 white paper “Building a Trusted Ecosystem for Millions of Apps,” https://www.apple.com/privacy/docs/Building_a_Trusted_Ecosystem_for_Millions_of_Apps.pdf.

[56] In the IBM antitrust proceedings (the case spanned 1969 to 1982), IBM responded at one point by dumping 30 million pages of unsorted documents at the steps of the courtroom, thinking that this would overwhelm the plaintiffs. They did not anticipate that one of those plaintiffs—Control Data Corporation—would fund an effort to scan and index the entire corpus, thus empowering their side of the litigation. See https://books.google.it/books?id=GAVOAwAAQBAJ&pg=PT187&lpg=PT187&dq=ibm+fought+antitrust+case+with+millions+of+documents&source=bl&ots=PpsY5PKL8e&sig=ACfU3U2bewMU3-aimC2Q8pN-npgBUJTkGQ&hl=en&sa=X&ved=2ahUKEwirkI7sv_LxAhXMyaQKHTZoDCAQ6AEwCHoECBsQAw#v=onepage&q=ibm%20fought%20antitrust%20case%20with%20millions%20of%20documents&f=false.

[57] L. Lessig, *Code: and Other Laws of Cyberspace* (Basic Books, 1999). Lessig is an American academic, attorney, and political activist and the Roy L. Furman Professor of Law at Harvard Law School.



Technology and Trust

The Big Picture

Trust is a major transaction cost. Indeed, as other costs recede in a frictionless economy, the ***dominant*** transaction cost.

When the stakes are sufficiently high, the transaction value justifies whatever technology is available to enable the transaction. Distrust is managed by minimizing ambiguity and ultimately, if necessary, exercising legal Authority (through negotiations, lawyers, contracts, litigation, whatever, ultimately reducing the trustor's risk). If digitization has any role, it is subordinate, as an operational improvement.

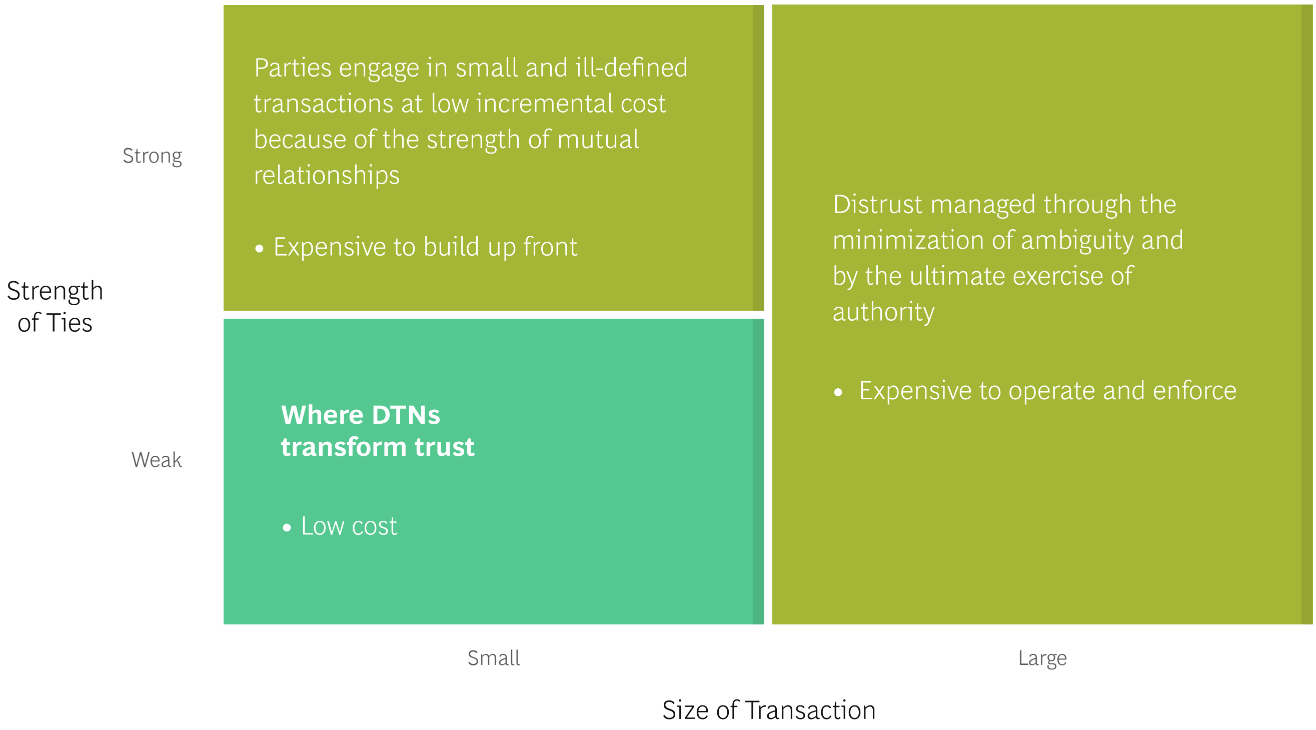
When ties are strong, parties can engage in even small and casual transactions at low cost because of the strength of their mutual Relationships. People exchange mutual favors without keeping account. Technology may allow more frequent reconnections and perhaps sustain those strong ties over a longer period of physical separation, but its benefit is marginal. Consider Toyota, which has shown the world that greater value can be created by nurturing trust in corporate Relationships than by managing distrust by appealing to legal Authority. But the story of the Aisin Seki fire, described on page 9, was a low-tech story; had Jive or Slack or Google Docs been available as collaboration platforms, the response would have been essentially the same, perhaps a bit quicker. Again, digitization would merely be an operational improvement.

So, for individuals as well as corporations, digitally intermediated trust is a marginal addition when the transaction is large or the ties are already strong. Conversely, it is transformative when the ties are *weak* and the transactions are *small*. And that, of course, describes the vast majority of transactions in the modern economy. (See Exhibit 3.)

Some bases of trust can be established for small transactions among parties with weak ties, while others cannot. Specifically: Reputation-at-Risk, when mediated in a DTN, can be massively scaled and indeed exhibits strongly *increasing* returns; Shared Identity can be scaled to bind together intense niche communities; and we are seeing signs of how Empathy will be extended in the future with the advent of augmented reality, wearables, and the like, as they heighten the sense of co-location in a shared context, between people at a distance, and even between people and machines.

Technology also transforms the methods of managing *distrust*. Data Control in a DTN (monitoring, transparency) becomes a lot more powerful but also controversial (invasion of privacy). Constraint becomes more granular, consistent, and instantaneous as technology enables minute control over parties' behavior. When employed together, Data Control and Constraint are mutually reinforcing and characterized by continuous, exponential improvement. Authority (exercised by bosses, judges, and so on), in contrast, is mainly a human process of adjudication and so is not much enhanced by the tools of technology. Thus, the big shift in how distrust is being managed is that cheap digital systems that limit untrustworthy behavior *ex ante* are replacing expensive human systems that reward and punish behavior *ex post*.

Exhibit 3 - Where DTNs Are Transformative



Strength of Ties

Strong

Weak

Parties engage in small and ill-defined transactions at low incremental cost because of the strength of mutual relationships

- Expensive to build up front

Where DTNs transform trust

- Low cost

Distrust managed through the minimization of ambiguity and by the ultimate exercise of authority

- Expensive to operate and enforce

Small

Large

Size of Transaction

Source: BCG Henderson Institute analysis.

With respect to both generating trust and managing distrust, new intermediaries are needed for digital trust networks. Even when the DTN is centered on a blockchain, so-called "oracles" are needed to generate and publish real-world information that initiates pre-programmed transactions. With respect to such trust-related matters as the provenance of a good or the labor conditions of its farmers or miners, there is a dearth of trusted sources of certification. More generally, consumers and investors are increasingly concerned with the ESG profile of a corporation or product, but there are few metrics and trusted institutions to monitor, aggregate, and disseminate such information. The problem of business models for such enterprises is yet to be solved.

In addition, existing intermediaries need to extend their services to ensure or enforce trust. Opportunistic parties are continually finding new ways to game existing trust mechanisms. Social network platforms dealing with viral misinformation or retail platforms dealing with fake reviews are engaged in an escalating battle to preserve counterparty trust within their DTNs. Airbnb has found that to manage trust between travelers and hosts, digitized Reputation-at-Risk is insufficient: they also need explicit policies and sanctions, which can be controversial and expensive to administer.

Intermediaries, of course, need to be trusted by the parties using them. Other things being equal, it is in intermediaries' business interests to be trusted, and they have at their disposal precisely the same seven trust mechanisms. Stack architectures, in general, drive the "polarization of scale advantage" by which competitive advantage shifts to the very large (typically operating at the bottom of the stack) and the very small (typically operating at the top)—both gaining at the expense of medium-sized enterprises.[58] To a great extent, this same phenomenon takes place in the DTN stack: technology extends the ability of small parties, linked by weak ties, to consummate trusting transactions. But increasingly, this requires that those parties, in turn, trust large enabling intermediaries, which enjoy powerfully increasing returns (stemming from the amortization of fixed development costs as well as from network effects). However, these enabling intermediaries are not, as perhaps they would wish, neutral enablers of the "wisdom of crowds." Rather, they are perforce trustees in their own right—trustees, moreover, that enjoy an unprecedented measure of market power. Specialist entrusted intermediaries have long been bound by moral and legal constraints as fiduciaries. Increasingly, these new trust platforms need to be managed with the same philosophy.

[58] This argument is made, for example, by Philip Evans and Patrick Forth in "Borges' Map: Navigating a World of Digital Disruption," BCG essay, April 2015. https://www.bcg.com/publications/2015/borges-map-navigating-world-digital-disruption. See also Evans's 2013 TED talk, "How Data Will Transform Business," https://www.ted.com/talks/philip_evans_how_data_will_transform_business.

Digitally intermediated trust is transformative when ties are weak and transactions are small. And that describes the vast majority of transactions today.

A Blueprint for Unlocking Value

As we have emphasized, transaction costs are the principal shaper of industrial organizations and have become a paramount constraint on economic performance. And as other transaction costs have receded, trust has emerged as the last obstacle, largely immune to technological fixes. The blockchain "movement" is a self-conscious and frontal attack on this challenge, but blockchains address only part of the problem: the trusted curation of a database that intermediates transactions.

The larger systemic context is the digital trust network, and it is the overall redesign of a DTN that unlocks value. The DTN stack serves as a blueprint for designing better systems for trust.

Every DTN is unique—unique with respect to its requirements for trust among the parties and also with respect to the combination of mechanisms currently implemented to generate trust or manage distrust. Thus, there is no universal template or off-the-shelf solution. Each DTN needs to be reviewed afresh. In that regard, the challenge of building a digital trust network is strikingly similar to that posed by process reengineering two decades ago: the elements are quite straightforward, but the value is created from a rigorously end-to-end implementation.

The starting point is a top-down review of the current trust network. This consists of four preparatory steps that are close to the method of process reengineering:

- **Map** how trust or distrust at each key interface facilitates or impedes the end-to-end performance of the business system. The challenge is to assess trust from the point of view of each party and thus achieve a neutral understanding of how trust affects the performance of the whole.

- **Conduct a gap analysis** to identify what benefits could be achieved with "perfect collaboration," the dollar cost of current trust management mechanisms, and possibly the ability of the system to scale.

- **Focus on the most serious gaps** and identify the bases of trust and the current mechanisms for curating that trust or managing distrust. The two frameworks offered in this paper may facilitate this analysis and provide a distinct contribution to better design systems for trust.

- **Develop a technology roadmap** by planning how the technologies specific to each mechanism enumerated above can be applied to increase trust or reduce the cost of trust management.

When control of a DTN is in the hands of a single orchestrator, it can proceed to develop and implement a DTN design. However, in many contexts, no one party sets the rules unilaterally, and so the diagnostic needs to be shared and some measure of consensus developed.

To be sure, the politics can be harder than the technology. But quite often, the most powerful incentive for transacting parties to tackle the failings of their common DTN is the knowledge that if they do not, some disruptor will do it to them!

About the Authors

Philip Evans is a senior advisor to BCG and a BCG Fellow. He is the author of many publications and is a coauthor with BCG's Tom Wurster of the best-selling book ***Blown to Bits***. You may reach him at evans.philip@advisor.bcg.com.

Matt Williams is a consultant in the firm's Washington, DC office and an ambassador with the BCG Henderson Institute. You may contact him by email at williams.matthew@bcg.com.

Marcos Aguiar is a managing director and senior partner in the São Paulo office of Boston Consulting Group and a fellow of the BCG Henderson Institute. You may contact him by email at aguiar.marcus@bcg.com.

Santino Lacanna is a principal in the firm's São Paulo office and an ambassador with the BCG Henderson Institute. You may contact him by email at lacanna.santino@bcg.com.

For Further Contact

If you would like to discuss this report, please contact the authors.