Decentralized technologies, governance, and institutional embeddedness

Bodó, B.

Publication date
2018

Citation for published version (APA):
In this paper, I’ll briefly discuss three modes of establishing the trustworthiness of trust mediators: technological systems which we use to produce and mediate interpersonal and institutional forms of trust. The first approach builds trust through creating objectively measurable systems. The second mode is institutional, and relates to the embeddedness of trust mediators in formal and informal institutional networks of trust production. The last approach is that of regulation. Due to space constraints I’ll only discuss them in a detail that is necessary to point out the relative merits and limitations of each approach.

1.1 Trust created by standardized objectification

The blockchain technology proponents claim that the distributed ledgers will be able to produce trust in themselves, and between others because the technical architecture ensures that opportunities for external (i.e state) and internal coercion, and control will be limited; the code based ordering regime is transparent, accountable, and immune to cheating. Decentralization removes points of coercion and control. Transparency of the ledger improves accountability. The cryptographic immutability of the records prevents cheating. The self-enforcing logics of smart contracts prevent the arbitrary application of rules.

Indeed, ledgers (shared or not), are, and have been since the 17th century, the instruments of standardized objectification, the fundamental component of modern power. A shared, cryptography based database is not just a set of records known to all, but a tool for social coordination and ordering (DuPont 2014). As Porter (1996) eloquently demonstrates, quantification, objectification and standardization are technologies to deal with distance and distrust1. From the 18th century onwards, using the scientific methods of natural scientists, bureaucrats in business and in public administration started to build their own tools to construct an “objective” representation of the underlying social, economic, geographic, and demographic realities, and developed the decision supporting argumentative instrumentations that converted these “objective” statistics into disinterested, efficient and optimal (rather than intuitive, value

---

1 “The rigor and uniformity of quantitative technique often nearly disappear in relatively private or informal settings. In public and scientific uses, though, mathematics (even more, perhaps, than law) has long been almost synonymous with rigor and universality. Since the rules for collecting and manipulating numbers are widely shared, they can easily be transported across oceans and continents and used to coordinate activities or settle disputes. Perhaps most crucially, reliance on numbers and quantitative manipulation minimizes the need for intimate knowledge and personal trust. Quantification is well suited for communication that goes beyond the boundaries of locality and community. A highly disciplined discourse helps to produce knowledge independent of the particular people who make it.” (Porter 1996, ix)
laden, and unaccountable) decisions. The rapid success of natural sciences and engineering to provide reproducible, quantifiable, externally verifiable explanations, and working solutions to a wide range of natural phenomena, as well as practical challenges, created a demand to develop similar approaches in other domains as well. The challenge to quantify, compare, explain, predict, and ultimately govern the activities, lives (input, and output) of local populations, and economies, required the development of standardized methods of survey and measurement, quantitative (statistical) analysis, and impersonal reasoning. As the industrial revolution gave birth to a techno-social system in which humans and machines were producing each other’s inputs and outputs in the form of labor, raw materials and products, an increasing number of human dimensions and activities required a form of description and understanding that matched the engineering and economic categories by which machines were described: inputs, energy, throughput, production, efficiency, etc. (Beniger 1986) The private sector and the markets developed a wide range of technologies of standardization and quantification to facilitate production and trade: standardized units, measures, weights; standardized norms of production and labor in the form of Taylorism, and scientific management; standardized and easily quantifiable tasks at the assembly line. In parallel, business bureaucracies developed their accounting practices that were able to offer an “objective”, quantification based, and comparable view of the activities of the organization for the owners, managers, and regulators. Public administration faced similar challenges when it had to compare public investment demands / alternatives, and when had to fulfill its regulatory obligations, and oversee the activities of diverse market actors. Foucault (1991a) links this process of quantification to a systemic and fundamental shift in how power is imagined to operate, and argues that the beginning of this process marks the emergence of governmentality: “the ensemble formed by the institutions, procedures, analyses and reflections, the calculations and tactics that allow the exercise of this very specific albeit complex form of power, which has as its target population, as its principal form of knowledge political economy, and as its essential technical means apparatuses of security.”

Ledgers are part of a complex assemblage of processes and technologies that were developed to quantify previously unquantified phenomena, in a standardized, thus comparable manner, ready for verifiable, objective, and disinterested analysis, upon which predictions and decisions can be made. Ledgers don’t just record the quantified traces of human activities, but through what and how they measure, and through how and to what aim they are used, they form the backbone of modern bureaucratic control, and ultimately modern power relations (Foucault 1991b; 1991a; 1979).

Distributed ledgers, and self-enforcing smart contracts are the next, logical steps in that form of administrative power. They amount to an infrastructure that records and publishes facts that comply with the pre-defined rules in an impersonal, ‘objective’, verifiable manner, so other, impersonal, verifiable, objective agents (i.e smart contracts, DAOs, etc) can make impersonal, verifiable, objective decisions that are thus free from arbitrary, subjective, non-verifiable, non-transparent influences, biases, motives, in one word, corruption. This is the way with which blockchain technologies promise to produce trust in themselves, and among their users.

The element which guarantees the impersonality, verifiability, and ultimately the trust is the same thing that underwrote the objectification of governance since the 18th century: science. The laws of nature are

---

2 “Mechanical objectivity has been a favorite of positivist philosophers, and it has a powerful appeal to the wider public. It implies personal restraint. It means following the rules. Rules are a check on subjectivity: they should make it impossible for personal biases or preferences to affect the outcome of an investigation.” (Porter 1996, 4)
clear, rational, transparent, public, verifiable by any- and every one. The deductions that are made through their application are also clear and indisputable. Engineers, statisticians, economists, and blockchain technologists all base their claims to objectivity by relying on the indisputable, public facts of mathematics, and physics: “We don’t need third parties to provide trust any more, because we have math.” (Dingle 2018, 23), or “when it comes to money, only math can be trusted” (Wright and Filippi 2018, 205)

These characteristics, at least in theory, enable technological systems to establish their trustworthiness independent from the pre-existing institutional order. But that applies to open technical systems, which, through their transparency of code and governance enable independent external scrutiny and verification. For closed systems that is hardly an option, and they need to find other sources of trustworthiness.

1.2 Trust through institutional embeddedness

Shapiro (1987) argues, impersonal systems are trusted because they are embedded in a complex, and interdependent networks of institutional control and oversight, procedural norms, structural constraints, and insurance-like arrangements: a complex system of checks and balances. In that logic, guardians of trust guard each other. The embeddedness of technological systems in the local social, economic, political institutional arrangements means that these local institutions are able to exercise different forms of mostly informal power over trust mediating technologies. These powers can take widely different forms. Do the rules, protocols, norms and values enforced by digital technologies reflect the local values, customs, norms? Are the local regulations recognized, respected, and enforced? Does a global entity have a local address, office, representatives who are responsible for keeping in touch with local clients and customers? Are the locally generated revenues taxed? Is data harvested off locals locally stored and processed? Do markets punish firms’ illegal, or highly irresponsible behavior? Do markets provide insurance against the risks these technologies represent. To what extent different entities have the power to shape the nature of contractual arrangements to control the behavior of these entities?

There is a constant friction between the standards planetary scale technologies carry, and the values which are expressed in the local norms, and institutional routines, and the nebulous practices of local communities. The outcome of how these frictions are ultimately resolved depend on the power balance between trust mediator technologies, and local institutions / stakeholders. This power balance determines to what extent technologies that operate on a global scale, and depend on the automated standardized, and objectified organization of social, economic relations relationships need to take into account the differences in local conditions, and the non-standard expectations of different local stakeholders. To illustrate this point with probably the most straightforward example, consider how the internet was seen some time ago as an unstoppable vector of USA First Amendment inspired speech rules across the globe. This norm assumed that it is the government which poses the most ardent threat to free speech, and a vibrant marketplace of ideas is a direct precursor to open, democratic societies. With time, this utopian, optimistic approach had to face uncomfortable realities. Free speech may not automatically lead to a marketplace of ideas which effectively filters out trolls, hate speech, or misinformation. Not all democratic communities have the same history and approach to free speech, and some prefer to ban, for example, Nazis from the public discourse. Other governments don’t see an uncurtailed marketplace of
ideas as a desirable thing, and they do whatever it takes to control online discourse (Morozov 2012). It turned out, that it is both impossible and probably undesirable to force US type speech regulation on the heterogeneity across and within different localities, each with its own history, issues, power relations, dynamics. But taking these local norms into consideration is far from being straightforward. These norms are constantly expressed, debated, contested, reaffirmed in local contexts. The debates and the participants identify the external contours of the “local”. While, for example traditional online community managers, forum administrators and content moderators must unavoidably participate in debates around the speech norms on their properties, platforms, operating on a planetary scale often do not. It is more efficient to offload content moderation tasks to AI systems, or underpaid and overstressed overseas subcontractors, isolated from the communities they are set to police, based on guidelines set in independently from the speech communities on which they are applied. (Newton 2019b; 2019a)

The reasons for the weak embeddedness may be economic as much as political. The economic factors are mostly related to the fact that the technical components are much easier to scale than those which involve humans, and their messy dilemmas. The scalability of technical components produces standardization, while the incorporation of diverse local norms and values creates frictions, uncertainty, and is costly. This explains why these technological systems are only weakly embeddedness in, in not outright hostile towards these existing institutional frameworks of informal control and oversight (Yeung 2017; 2019).

Embeddedness mostly relies on voluntary compliance by technology operators, and takes the form of self-, or co-regulation. If that does not take place by itself, regulation must step in and coerce technologies to comply with a minimum set of standards.

1.3 Regulation trust mediating technologies

The debate on the relationship between trust and regulation, especially in the US discourse, boils down to the differences in opinion whether regulation destroys or crowds out trust. Regulation skeptics argue that regulation, both in legal (Ribstein 2001) and in technological (Nissenbaum 2001) forms may have the blowback effect of destroying naturally emerging trust by, for example, signaling untrustworthiness, or replacing trust with increased control. Those who don’t believe the effectiveness of bottom-up, market based, or other, decentralized, informal forms of trust formation, advocate for various forms of regulatory frameworks to support, if not replace bottom-up trust. (see (Balkin 2016), and footnote 79 in (Hall 2002)). In Europe, where there is less traditional resistance to regulation, trust, and other, informal logics of ordering play a less prominent role in the debates on the necessity of regulation.

In that debate, Hall (2002) suggests that when trust in not purely instrumental, but has significant interpersonal relevance, regulation can interact with trust in three distinct manner. “In its predicated attitude, the law takes the existence of trust as a factual premise for imposing a particular rule. [...] The legal rule does not depend on any assumptions about how the law affects trust, only that there is trust. Trust is the source of law, not its object. In contrast, the supportive attitude arises from attempts to use the law to increase or sustain trust. In this second category, law produces trust, rather than trust producing law. Third, the skeptical legal attitude about trust uses the absence or illegitimacy of trust as a premise for a legal regime that institutionalizes distrust.” (Hall 2002, 486)
The first attitude takes trust as given. The regulation is crafted around pre-existing trust, with the aim to create incentives for the trustee to maintain their trustworthiness. Fiduciary law and informed consent obligations play a role in both healthcare law, which Hall has originally considered, and the regulation of trust mediating technologies. In the supportive logic, the aim of regulation is to directly or indirectly foster trust. This approach recognizes both the fact that trust may be contingent on certain safeguards, backstops, confidence inducing legal frameworks, and the fact that in some situations their creation or strengthening might be necessary to enable the development or maintenance of trust. Hall suggests confidentiality obligations/protections in the healthcare context which has the structural equivalent of privacy protection in the information law domain. Finally the skeptical attitude aims to address situations where distrust needs to be mitigated, or where trust is misplaced, and needs to be rebased, or when one must expect ongoing need for collaboration even in the absence of trust because of some form of power asymmetry. The range of situations may vary widely from situations when regulation may help distrustful parties to take the leap of faith, to the other extreme, where formal arrangements are necessary to address strong and persistent conflicts of interest.

As we have seen before, different assumptions about the current level, and foreseeable dynamics on pre-existing trust will have an impact on the choice of regulatory instrument, and their effectiveness. A very different regulatory approach follows from the assumption that trust in trust mediators still exists, compared to an assumption that distrust is not just unavoidable, but a healthy degree is even desirable. In the next three subsections I’ll briefly assess different regulatory approach from this perspective.

1.3.1 Can trust be taken for granted?
The regulatory approach that assumes the existence of trust vis-à-vis regarding trust mediators focuses on the concept of information fiduciaries. First raised by Balkin (2016) in the US context, its starting point is the following: “Generally speaking, a fiduciary is one who has special obligations of loyalty and trustworthiness toward another person. The fiduciary must take care to act in the interests of the other person, who is sometimes called the principal, the beneficiary, or the client. The client puts their trust or confidence in the fiduciary, and the fiduciary has a duty not to betray that trust or confidence. [...] Fiduciaries have two basic duties. The first is a duty of care. [...] The second, and in many ways more important duty, is the duty of loyalty. Fiduciaries must keep their clients’ interests in mind and act in their clients’ interests. As a result, fiduciaries also have a duty not to create potential or actual conflicts of interest that might undermine their duties of loyalty and lead them to undermine the interests of their clients.” (Balkin 2016, 1207)

This approach encapsulates an important implicit assumption, namely that that the interests of the technology and their users are or at least can be aligned, and the task is to provide incentives for technology operators to stay honest. I doubt that this can be taken as granted. On the contrary, I’d like to argue that with digital technologies, the interests of the technology and the users are fundamentally misaligned. In order for trust mediators to be able to fulfill that role, they need to engage in practices, such as surveillance, algorithmic ordering, which are not just very difficult to do in a trustworthy manner, but also encode systemic conflicts of interest for the trust mediators. Zuboff (2019) argues that the whole

---

3 It reflects many of the US specific concerns about regulation: the general mistrust in government, the primacy of strong constitutional safeguards protecting speech, and the absence (or political impossibility) of sector or domain specific regulation, akin to the European data protection regulation; and the specific constraints posed by case law, especially around commercial speech.
technology based economic regime by its own ruthless, instrumentarian logic must rely on total surveillance and the corresponding algorithmic adjustment of behavior. Gürses and Van Hoboken (2018) make a similar point about the technical aspects of these systems develop, when they argue that the pervasive surveillance is not just the ultimate outcome, or goal of these technologies, but a critical input resource, without which current logics of technology development cannot operate. This means that on the one hand trust mediators must ensure the total transparency of their users, and remain totally intransparent to prevent the gaming of their systems, and protect themselves from distrust, and maintain their economic, social position as trust mediators. In addition, users are not the most important stakeholders in the complex web of interests trust mediating technologies are situated in. The shareholders', investors' profit expectations; the sellers' and advertisers' expectations to be able to better control user behavior; the interest of governments to enforce the law; and the interests of those who wish to have access to users' behavioral surplus data are all in direct conflict with the legitimate interests of users (Zuboff 2019). This means that when it comes to privacy, when it comes to fairness in algorithmic decisions, when it comes to minimizing profiling, when it comes to individuals' preferences, as well as fundamental rights to how they want to be seen by, and represented on, and served by trust mediating technological systems, there are fundamental conflicts of interest, which makes it very difficult to design regulation which is based on the encapsulated interests, or in the belief in the good faith motives of the trust mediators.

1.3.2 Regulation to foster trust: the supportive logic
The supportive logic still operates under the conditions of trust, and aims to strengthen the individual and institutional frameworks around trusting behavior. To achieve that, it can address the substantial power- and information asymmetries between users and technologies (Bodó et al. 2017), or strengthen the ability of users to better exercise contractual, or commercial forms of control. (Richards and Hartzog 2015, 444) The justification for such regulatory intervention may be that with higher degrees of trust, users are more likely to use trustworthy services (if they have the choice), and less likely self-censor, or share misinformation, incorrect or insufficient information. (Richards and Hartzog 2015) The supportive logic focuses of the trustor, that is the user of trust mediating technologies. As such, it operates from the assumption that it is the individual that needs to be empowered to facilitate the development and maintenance of a trusting relationship. Regulatory proposals, such as the explainability of machine based decisions (Pasquale 2017), information disclosures on the fact and origins of automated recommendations and decisions, mechanisms that empower users in regard to the scope and amount of personal data collected, how it is handled and used in decisions; the ability to challenge decisions; the imposition of public duties (Balkin 2017); and strict liability regimes for technology operators are key examples in this domain.

1.3.3 Regulating under the conditions of distrust
The most aggressive approach to regulations starts from the assumption of distrust. For instance, this was the approach authoritarian regimes took towards western technologies, which they saw as a threat. Consequently such regimes were the first to exercise total and uncompromising control over digital

---

4 Mediated trust is like laws which are like sausages. It’s better not to see them being made.
technologies. By all accounts democratic societies seem to also have arrived to a point of juncture regarding the trustworthiness of trust mediating technologies, as the very trustworthiness of these technologies are increasingly called into question.\textsuperscript{5} There are growing concerns that trust mediators do not possess the necessary competence, benevolence and integrity to fulfill expectations both context specific and generic. There are many different concerns around the trustworthiness of trust mediators. Their inability to detect and control foreign government meddling with domestic democratic processes turned mere trustworthiness into a hard-core national security issue. Their dominant positions in multiple key markets, such as retail, social media, news, turned size into both a competition issue, and a worry about allowing companies to grow too big to fail\textsuperscript{6}. End-to-end encryption of communications threatens state law enforcement powers by creating de-facto extrajudicial territories.

These concerns can be used to justify regulation that is guided by the conditions of mistrust, rather than trust. In fact we already see some regulation put in place by authoritarian regimes, when they completely block certain technologies in times of crises, require technology operators to collaborate fully with local authorities, or force them to censor some of the interactions taking place among users. In democratic societies the stakes are no less contentious, and regulatory momentum is clearly driven by new conditions of distrust.

\section{Conclusions}

We live in a technological environment where our interpersonal and institutional logics of trust production are mediated by digital technologies that operate on a scale well beyond the traditional action radius of pre-existing trust production frameworks. This, as well as their scale, speed, fluidity, and their weak embeddedness into the existing trust production frameworks puts trust mediating technologies in the curious position, where their trustworthiness is in question. For a long while the question of why we should trust trust-mediators could be bypassed because these technologies enjoyed strong ideological support about their beneficial role in society, and because it was assumed that their interests are well

\footnotesize\textsuperscript{5} Puppis and Winseck (2019) curates a growing list of technology regulation proposals from around the globe, listing more than 50 proposals in August 2019.

aligned with those of their users. We almost reached a point, where trust mediators seamlessly blended into the background, and became an invisible infrastructural element in our wider networked societies. Yet, in the second half of the 2010’s a series of serious crises pointed to the fact, that we have few reasons to be confident in the competence, benevolence and integrity of technological trust mediators. As a result, a flood of legislative proposals are being put forward in all major jurisdictions that try to address among others, data protection and privacy concerns, the transparency, accountability and fairness of machine learning decisions, the price setting practices of online marketplaces, the concerns around user profiling, including facial recognition, the content moderation practices of various digital intermediaries.

The different regulatory proposals directly or indirectly will affect the trust in these technologies. They may start from different assumptions regarding the pre-existing conditions of trust and distrust these systems operate or should be operating. This paper drew together various strands of the trust literature to provide the first steps towards a comprehensive trust assessment framework. The ultimate goal is to connect two, currently distant, and only indirectly connected challenges. On the one end we have billions and billions of human beings who use digital technologies, and that ultimately brings changes to how they develop trust among each other. On the other end we have highly complex techno-social systems which directly and indirectly mediate trust, and in the same time need to establish their own trustworthiness in the myriad contexts in which they are used. This is in part a hard regulatory question. But it has important technological, architectural aspects, as well as raise the question of how we can embed these technical systems in local trust architectures, which include communities, their institutions, norms, practices, routines, systems of familiarities.

This task requires that we don’t let trust mediators sink into the background, at least for the time being. There may be a time when we can let them turn into trust infrastructures, and forget about them. But we need to know if we can trust them first.

3 References
