Blockchain Technology, Trust & Confidence: Reinterpreting Trust in a Trustless system?

Primavera de Filippi, Morshed Mannan, Wessel Reijers, Paula Berman, Paula Berman, Jack Henderson

▶ **To cite this version:**

Primavera de Filippi, Morshed Mannan, Wessel Reijers, Paula Berman, Paula Berman, et al.. Blockchain Technology, Trust & Confidence: Reinterpreting Trust in a Trustless system?. Humboldt Institute for Internet and Society. 2022. hal-03895402

HAL Id: hal-03895402
https://hal.science/hal-03895402

Submitted on 12 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

PRIMAVERA DE FILIPPI, MORSHED MANNAN, WESSEL REIJERS, PAULA BERMAN, JACK HENDERSON

Blockchain Technology, Trust & Confidence

Reinterpreting Trust in a Trustless system?

HIIG DISCUSSION PAPER SERIES 2022-3

ABSTRACT

This report provides an in-depth analysis of the theoretical foundations of the concepts of trust and confidence and their correlation to the notions of risk, agency and legitimacy. These theoretical underpinnings are thereafter applied to permissionless blockchains, in order to identify how the foundational nature of the technology lends itself to bridging traditional paradigms in generating greater trust and confidence amongst users. We first review the literature which distinguishes the concept of trust from confidence before turning to major scholarly contributions on interpersonal trust, system trust and trust in technology. This literature has been examined through the lens of a multi-disciplinary reading group, which subsequently brings to bear practical and philosophical considerations in the use of blockchain technology in the modern world. The report concludes by providing the key areas in which trust can be reinforced, identifies the limitations of traditional factors as a measure for building trust with a decentralised and pseudonymous environment and clearly defines the differing levels of trust and confidence which are required within an on-chain and off-chain governance structure and its nexus to engendering legitimacy.

KEYWORDS

Trust, Confidence, Legitimacy, Blockchain, Technology

CITATION

De Filippi, P., Mannan, M., Reijers, W., Berman, P. & Henderson, J. (2022). Blockchain Technology, Trust & Confidence. Reinterpreting Trust in a Trustless system?. HIIG Discussion Paper Series 2022-3. 20 pages. https://doi.org/10.5281/zenodo.6516991.

LICENCE

This work is distributed under the terms of the Creative Commons Attribution 4.0 Licence (International) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (https://creativecommons.org/licenses/by/4.0/). Copyright remains with the authors.

AUTHOR INFO / AFFILIATION

Primavera De Filippi is Research Director at the National Center of Scientific Research in Paris, Faculty Associate at the Berkman-Klein Center for Internet & Society at Harvard, former member of the Global Future Council on Blockchain Technologies at the World Economic Forum, founder and coordinator of the U.N. Internet Governance Forum's dynamic coalitions on Blockchain Technology (COALA).

Morshed Mannan is a Max Weber Postdoctoral Fellow at the Robert Schuman Centre for Advanced Studies at the European University Institute in Florence and a Research Affiliate of the Institute for the Cooperative Digital Economy at The New School.

Wessel Reijers is a postdoctoral researcher at the Department of Philosophy at the University of Vienna, and Visiting Scholar at the Robert Schuman Centre, European University in Florence.

Paula Berman is a democracy researcher, civic technologist and COO at RadicalxChange Foundation.

Jack Henderson is a digital democracy researcher with RadicalxChange Foundation, the ERC's BlockchainGov project, and the Coalition Of Automated Legal Applications. Jack holds a degree in economics from Princeton University.

FUNDING INFORMATION

The report was funded by the European Research Council, Grant Agreement No. 865856.

CONTENTS

1 INTRODUCTION

Blockchain technology is often described as a revolutionary technology that enables us to bypass traditional centralised intermediaries, replacing them with a system based on mathematics and cryptographic proof. "*Vires in numeris*" (latin for "strength in numbers") has become one of the mottoes of early blockchain communities, advocating for the novel characteristics of this technology, in terms of decentralisation, tamper-resistance, transparency, verifiability, and—most importantly—trustlessness (i.e., the idea that, as long as we trust the underlying technology, we do not need to trust anyone else). Today, the technology has reached mainstream adoption, especially with the advent of decentralised finance applications, with a transfer volume of over 15 trillions dollars for Bitcoin alone in 2021.[1] Yet, most people do not necessarily understand the underlying complexities of such technology, and blindly "trust" these infrastructure to be, indeed, "trustless".

Commonly, it is assumed that blockchain technology was created as a response to the trust crisis that swept the world in the wake of the 2008 financial crisis. Bitcoin and other permissionless blockchains were presented as a "trustless" alternative to existing financial institutions and governments. The trustless nature of blockchain technology has been heavily questioned, but little research has been done as to what blockchain technology actually brings to the table in the place of trust. Drawing from this distinction, a recent article by Primavera De Filippi, Morshed Mannan, and Wessel Reijers (De Filippi et al., 2020) argues that blockchain technology is not a trustless technology but rather a *confidence machine*.

In his seminal paper "Familiarity, Confidence, Trust: Problems and Alternatives" (2000), Niklas Luhmann argues that both trust and confidence are about expectations. In a situation of trust, there is a perceived risk that one's expectations will be disappointed, but one freely chooses to trust anyway, making oneself *vulnerable. Ceteris paribus*, the decision to trust will depend both on the amount of trust that can be conferred and the possible consequences that a breach of trust might entail. A situation of confidence is characterised by the *lack* of perceived risk and vulnerability. The person is confident that their expectations will not be disappointed.

Blockchain technology deploys cryptographic rules, mathematics, and game-theoretical incentives to increase confidence in the operations of a computational system. Blockchain-based networks, thus, produce confidence not by eliminating trust but rather by maximising the degree of confidence in the system to indirectly reduce the need for trust. A higher degree of confidence allows transactions to occur more easily by reducing the perceptions of risk associated with these transactions. However, such an increase in confidence ultimately relies on the proper operation and governance of the blockchain-based network and it is affected by the underlying Internet-level governance and operation (De Filippi & McMullen, 2018).

This leads us to an important distinction regarding the dual governance system of blockchain networks: the enforcement of rules generally happens *on-chain* (i.e. governance *by* the infrastructure), but the procedures to elaborate and eventually change these rules are usually done *off-chain* (i.e. governance *of* the infrastructure). On-chain governance is generally achieved by incorporating technological guarantees directly into the blockchain network's technical fabric and thus building confidence. Off-chain governance, on the other hand, cannot be achieved by technological means alone, and consequently requires elements of trust. In other words, it requires trusting a variety of actors, including nodes, miners or validators, developers, all the way to third-party software applications and institutional actors—thereby

[1] https://www.nasdaq.com/articles/bitcoin-transfer-volume-now-exceeds-%2415.8-trillion-2021-07-24

creating a *concentric* relationship between trust and confidence.

The aim of this report is to analyse the relationship between trust and confidence in blockchain-based systems and foster future discussions on how different levels of trust and confidence, both on-chain and off-chain, can influence the perceived degree of legitimacy of blockchain-based networks. The report is intended to assist academics and researchers that are investigating the concept of trust to deploy existing theories on trust (and allied concepts) to emerging technologies, such as blockchain, so as to build and refine these theories. While the report synthesises key texts on the topic, and evaluates this body of literature in light of its implications for understanding blockchain technology, its objective is not to be purely of theoretical interest. Instead, the report connects theory to practise, as it incorporates reflections from blockchain communities who engaged with these texts, as well as insights from their practical experiences. In turn, the report seeks to acquaint developers, lawyers, and other stakeholders with these important topics, and help them develop a critical perspective on the broader narrative of blockchain as a 'trustless' (a.k.a confidence) machine. It is to this end that the report provides an analytical overview of the readings and discussions that were part of the multidisciplinary BlockchainGov reading group on "Trust, Confidence, and Blockchain Technology." The participants of the reading group ranged from social science researchers to lawyers to software developers to philosophers.

Launching a reading group on trust is a daunting task as the literature on the topic is voluminous, so we adopted a purposive sampling method rather than attempting to be exhaustive. The premise of the reading group was to explore the distinction between the concepts of trust and confidence in the context of a particular technology, so the literature was purposefully selected on the basis of how they contribute to theorising this distinction and how these concepts (along with allied concepts of faith, reliance and security) were applied to the study of technological systems. Note that the order in which the papers are discussed in this report are not in the order in which they were read.

Luhmann's aforementioned paper provided our starting point. The chapters by Gambetta and Hardin were selected as they are in conversation with Luhmann's work and discuss the delineation of confidence from trust, while making their own distinct interventions. The paper of Pettit, and the more recent article and chapter by Nguyen and Donath engages extensively with aspects of what Luhmann terms 'confidence', such as the existence of an "unquestioning attitude" and relations based on "active reliance" or calculations of low-risk (i.e., not "interactive reliance" or "affective trust"), even when not entering into a discussion of his work specifically. Relatedly, empirical studies have also begun to specifically focus on the concept of confidence, and the interplay between trust and confidence, when studying institutions like the police and voluntary organisations. The articles of Jackson & Sunshine (2007) and Tonkiss & Passey (1999) respectively provide select, highly-cited examples of such an approach, which contributes to a more nuanced understanding of what makes confidence different. This purposive selective sampling of literature was productive as it not only explored the distinction we were concerned with, but also provided a state-of-the-art overview of relevant discussions on interpersonal trust.

The second set of readings discussed how the concept of trust could be applied to institutions and (technological) systems. The scholarly work of Sumpf and Nissenbaum were important in this respect as they acted as a bridge from earlier theoretical literature on interpersonal trust to system trust and addressed a particular concern of system trust: security. This was complemented by the work of Simon and Coeckelbergh, who respectively wrote about trust in the Web and trust in robots, as they specifically positioned the system trust literature within technical domains that are closely related to blockchain: the Internet and automation technology. Finally, the chapter and recent article by Schneier and Bodo provide an outlook towards the future: knowing what we now know about trust (and confidence), how can we understand how trust is being mediated online at present and how can we build societal pressures to create

or promote trust?

The reading group focused on deliberating upon the following key considerations:

- Can trust in one system contribute to more confidence in another system? Can building trust in off-chain governance generate more confidence in the operations of a blockchain network?
- Can confidence in a system contribute to more trust in another system? Can on-chain governance build more trust in institutions adopting blockchain technology?
- How can we design blockchain-based systems that build confidence and, in doing so, create the conditions for trust to emerge off-chain?

The report is structured as follows. The first section elaborates on different definitions of trust, analysing the strengths and weaknesses of each conceptualization. It considers multiple conceptualizations of trust: trust as a rational calculation, trust as encapsulated interests, trust as reliance, phenomenological-social trust, trust as an unquestioning attitude, trust as security, and affective trust. The following section investigates the various factors that may contribute to increasing or decreasing trust and confidence: restraints on the agency of trustees, the establishment of trust mediators, socio-technical systems, the collective production and assimilation of knowledge, societal pressures, and institutionalisation. The final section concludes with a discussion of the importance of harnessing the benefits of both trust and confidence in blockchains so that risk is diminished and the blockchain operates securely while still leaving room for agency and freedom for action. Such a concentric and complementary relationship can help ensure the legitimacy of blockchain-based systems.

2 CONCEPTUALIZATIONS OF TRUST

Trust is a multifaceted concept that has generated extensive scholarly debate and a voluminous body of literature. For our investigation, this report does not attempt to provide a full account of its different existing conceptualizations. Instead, it focuses on those that help illuminate a fuller understanding of how trust relates to blockchain technology. The following section will present selected definitions of trust, in terms both "positive" (what trust is) and "negative" (what trust is not), and analyse the strengths and weaknesses of each.

2.1 Trust as a Rational Calculation

The conceptualization of "trust as a rational calculation" comes out prominently in Diego Gambetta's "Can we trust trust'?" (2000), where the author contributes to the existing literature on trust by exploring its relationship with cooperation. Trust here is defined as a threshold point located on a probabilistic distribution of subjective expectations, centred around a midpoint of uncertainty with complete distrust on one end and complete trust on the other.

Gambetta outlines three conditions under which trust becomes relevant for fostering cooperation: uncertainty, the agency of the trustor, and the agency of the trustee. Trust entails uncertainty, as it is ultimately a way of coping with information asymmetry; with unlimited computational ability to map all possibilities, trust might be unnecessary. Trust entails the agency of the trustor, for trust requires the freedom to choose to trust and make oneself vulnerable. Finally, trust entails the agency of the trustee: trust can only emerge if there is the possibility of *betrayal*. The trustee has the ability to betray the other, but still freely chooses to honour the trust. Without these conditions and freedoms, there can only be Luhmann's sense of confidence.

In Gambetta's view, trust can be both a precondition and result of cooperation, claiming that it is possible

to rationally trust trust and distrust distrust. This means deliberately choosing a probabilistic value of trust—high enough to engage in tentative action but small enough to lower the risk and scale of possible betrayal. In order to trust trust as a rational strategy, one must consider that there is no concrete evidence for "trustworthiness"—only a lack of proof for breach of trust.

Trust begins with keeping oneself open to evidence, acting as if one trusted, at least until more stable beliefs can be established based on further information. When trust is not unconditionally bestowed like blind faith, it may generate a greater sense of responsibility at the receiving end. In doing so, the concession of trust can generate the very behaviour which might logically seem to be its precondition.

Conversely, to distrust as a rational strategy, one must consider that distrust can provide its own evidence. It encourages us to look for evidence to disprove trust, creating a self-fulfilling prophecy where we become increasingly more distrustful. Being wrong must then be taken as an inevitable part of the wager between trust and betrayal, where only if we are prepared to endure the latter can we hope to enjoy the former. Asking too little of trust is just as ill-advised as asking too much, Gambetta warns.

Here, it is worth questioning whether trusting trust can be based purely on a rational motive or interest. That is, can one rationally choose to trust? How can we categorise collective intentions and beliefs if trust is merely a probabilistic distribution of subjective expectations? Such a game theoretical approach fails to address the variety of human rituals that foster trust with repeated pleasant interactions and voluntary cooperation. The coldness of game theoretical models may not fully capture the role of trust in our richly social lives.

2.2 Trust as Encapsulated Interests

In the first chapter of his book, *Trust & Trustworthiness* (2002), Russell Hardin elaborates on the actual—rather than just conceptual—ideas of trust by addressing how trust emerges in the real world, even in situations that don't necessarily encourage its emergence. As with Luhmann and Gambetta, Hardin's solution is game-theoretical and based on rationality.

Hardin defines trust as "encapsulated interest," whereby trust emerges from the trustor's expectation that the trustee has an interest in continuing the relationship with the trustor: "I trust you because I think it is in your interest to attend to my interest in the relevant matter" (Hardin 1). This means that trust is both functional (i.e., limited to some issues) and relational (i.e., limited to a particular person): A trusts B to do X. Hardin aims to encompass the full range of trusting relationships, from those based strictly on self-interest to those based on thick relationships and emotions.[2] Yet, Hardin believes that the most critical types of trust are those based on rationality and self-interest—whether these are monetary interests or forms of self-interests based on morality, reciprocity, or otherwise.

Viewing trust as encapsulated interest helps uncover the extent to which one might be rationally inclined to engage in a trusting relationship, recognizing that trust depends on the rationality of both the trustor and the trustee. Trust is ultimately derived from the trustor's rational assessment of the trustee's motivations.

However, this conceptualization is limited to the extent that it is grounded on a purely descriptive and transactional relationship between two parties, which remains independent of the motivational orientation of the trustee. Trust is maintained if the trustor believes that the trustee is genuinely concerned with her best interests, or if she believes that the trustee's actions are motivated by fear of retaliation, say, through the

[2] Hardin argues that "the sense that trust inherently requires more than reliance on the self-interest of the trusted may depend on particular kinds of interaction that, while interesting and even important, are not always of the greatest import in social theory or social life—although some of them are, such as the trust a child can have in a parent" (Hardin 2002, 7).

justice system in a case of unjustified defection.

Moreover, Hardin fails to account that trust is a much more complex phenomenon, with biological underpinnings, which may emerge as an emotional state rather than a strictly cognitive and rational process. As we will see, the act of trusting others and being trusted by others is not only instrumental to a particular end (e.g. facilitating the execution of a particular transaction). It can also be regarded as an end in and of itself, to the extent that experiencing trust constitutes a pleasant experience for both the trustor and the trustee on the receiving end.

2.3 Trust as Reliance

In "The Cunning of Trust" (1995), Philip Pettit describes a rational reason for trusting someone even when Gambetta's rational calculation might deem the trustee unlikely to be trustworthy. Pettit bases this reason on the fact that people are naturally "regard-seeking," and trust can serve as a signal that the trustee is well-regarded (212).

As long as the trustee sees a benefit in being regarded by the trustor as trustworthy, she will be inclined to act in such a way as to honour the trust she has been conferred. Displaying such trust in public where others can witness it will further increase her motivation to act as a reliable person. Pettit argues that the trustor can exploit this natural tendency to seek the esteem of others in order to strengthen the *reliance* of the trustee. Even without a belief in the person's pre-existing trustworthiness, the trustor can act as if that person was trustworthy, on the ground that such a display of trust will make them act reliably.

Pettit's interpretation of trust remains a game-theoretical and rational calculation, as opposed to an emotional or psychological one, but with the added layer of intersubjectivity. The interpretation does not focus as much on the trustworthiness of people but rather on their desire to be perceived as trustworthy, which might lead them to be reliable, even if not inherently trustworthy. It recognizes the inherently social context of trust, as the trustee's actions will depend not only on the intrinsic character of the trustee, but also on the desire to be regarded as reliable by others (which includes the opinion of the trustor and that of the more general public).

However, Pettit's analysis does not sufficiently acknowledge the potential connection between one's desire to be perceived as trustworthy and one's personality and relationship with trustworthiness. An individual who values trustworthiness as a virtue will desire to be perceived as trustworthy by others. Hence, not everyone will act as a reliable person upon the conference of trust, because there is an important correlation between the trustworthiness of people and their desire to be regarded as trustworthy by others.

At the same time, one's relationship with trustworthiness is not only an individual predisposition, but also a result of societal contingencies: how much one values a certain virtue depends on the context and community one finds themselves in. For instance, one might care less about being perceived as trustworthy in a social or institutional context that does not care about trustworthiness, than one where trust is considered an important virtue.

2.4 Phenomenological-Social Trust

Mark Coeckelbergh makes the case for a "phenomenological-social" (P-S) approach to understanding the emergence of trust, as opposed to a "contractarian-individualist" (C-I) approach, in "Can we trust robots?" (2012). Coeckelbergh considers the accounts of Luhmann and Gambetta to be C-I, which is the notion that trust emerges from atomized individuals who interact, establish relations, and build trust through those relations. It is a perspective that sees trust-building as a process that starts with individuals who then

socialise, producing trust relationships.

The P-S account counters that sociality comes first, situating people in preexisting trust relationships, and out of which individuals are emerging in the first place. Coeckelbergh's claim here is that trust is not produced, but rather embedded from the beginning within the social bonds of the community.

The P-S approach, which precludes the possibility of trustlessness so long as there are social relations, informs Coeckelbergh's understanding of trust in robots. He considers the question first by recognizing certain pre-conditions of trust: social relations, the ability to use language, and Gambetta's sense of the capacity for freedom and uncertainty, so that the trustee can misuse the given trust.

In the C-I approach, one would assess trust in a robot by asking if it does as intended or expected. Here robots, "more than mere tools," are agents alongside humans, with the capacity to trust and be trusted even in the absence of shared norms and values (57). Instead, the P-S approach asks whether the robot grew into a social relation and is felt as an other. What matters here is how robots appear rather than what they are: they may not have language nor freedom, yet we often treat them as if they did.

Until Coecklebergh's contribution, this report's definitions of trust and confidence—which are more limited and narrow than the broad territory they naturally occupy to reveal the borderline—would deem it impossible to trust a robot that lacks agency and thus the ability to betray trust. The trustor can trust the human developer or operator behind the robot, for they have agency and make the trustor vulnerable, but they can only have confidence that the robot will act as expected.

Coeckelbergh, however, reveals the subjectivity and affective dimension of people's relationships with objects, which can be so personalised and anthropomorphized that they are perceived to constantly honour the trust relationship. When a social bond like this develops, people can and do, indeed, trust robots, even if irrational.

Still, the feeling of trust in robots is usually better understood as confidence and trust in the humans behind it, so long as the robot lacks the freedom to betray trust. However, continued developments in artificial intelligence may deliver such freedom and make the question of trusting robots more relevant. In the meantime, we should be wary of robots designed and operated not for actual trustworthiness but only the impression of it—deceiving, manipulating and exploiting our instincts as social beings toward trusting.

2.5 Trust as an Unquestioning Attitude

Describing "Trust as an unquestioning attitude" (forthcoming), C. Thi Nguyen adopts Luhmann and Gambetta's language of vulnerability and the possibility for betrayal, but holds that this particular form of trust can be directed at not only people but also non-agential objects.

Most accounts of trust are based on the assumption that trust is oriented toward human agents who think about the interests of the trusting party and behave following those interests. Nguyen offers a way to think about trust that encompasses complex technical artefacts like devices and algorithms.

To take up an unquestioning attitude means to step away from a deliberative stance—to trust but no longer monitor, challenge or question the trusted. Nguyen sees this as the relevant form of trust when, for example, one uses a climbing rope: we suspend deliberation, often without realising, and take for granted its reliability. Others would say trust in objects is simply trust in the people behind them, but Nguyen points out that an object can depart from the manufacturer's circumstances and control. Upon buying a new rope, trust is indeed placed in the manufacturer, but the decision to use the same rope a few years later

is a more complicated form of trust.

In Nguyen's view, trusting with an unquestioning attitude is suspending deliberation, integrating and internalising the external world, and thus expanding our agency and cognition. When we attempt to "integrate [other objects] into our own agency, only to have them malfunction," Nguyen considers the psychological response to be a form of self-alienation and betrayal (5). In fact, he himself felt betrayed once when his computer misarranged his folders after being unplugged from a monitor; the rage he felt led him to consider these questions.

While trust may lead to this second-order act of unquestioning, there might be other reasons to reach this unquestioning attitude, and not all of them relate to trust. In Luhmann's sense of taking for granted that something will work in a predictable manner, confidence could also reach the point of unquestioning; we can be unquestioning that the sun will rise tomorrow, but out of confidence rather than trust. Familiarity and experience, knowledge and expertise could also reach this point.

The path to an unquestioning attitude might affect how we relate to broken expectations. In the case of broken confidence, we blame the external source, but when trusting expectations are broken, we blame ourselves for trusting and making ourselves vulnerable. Nguyen's point about internalising the external and blaming the self for its failure, placed within the framework of trust and confidence, might justify why we can feel betrayed by what went unquestioned.

2.6 Trust as Security

In her article "Securing Trust Online" (2001), Helen Nissenbaum defines the question of trust in information systems as comprising both insiders, such as system administrators, developers, owners of private platforms; and outsiders, which concern hackers or other types of external and adversarial agents.

Nissenbaum warns that a lack of internal security can inhibit trust. At the time of writing, the question of trust in information systems was especially relevant in e-commerce, where a vibrant digital economy could be hampered by the lack of trust, caused by digital fraud and hacks. Within that context, proposed solutions in more technical discussions aimed to achieve trust through security, as well as better access control, management and enforcement of identity, and surveillance.

While most of the focus was on creating secure systems to keep outsiders out, Nissenbaum saw that too little attention was placed on the question of trusting insiders, who might tamper with systems from the inside. The past decade has proven the timeliness of Nissenbaum's article. External security has been addressed for the most part, while the question of internal security that Nissenbaum raised is receiving more and more attention. Today, many online platforms have grown into monopolies, with extractive business models, unaccountable CEOs, lack of regulation, unexplainable machine learning models and other components that firmly assert the threat of the insiders as the primary challenge to the emergence of trust in information systems.

Despite the insufficient focus on internal security, the issue of excessive security crowding out trust remains, and Nissenbaum warns against the conflation of security with trust. She claims that security helps ensure confidence in Luhmann's sense, but at the cost of agency and vulnerability. As a consequence, we lose the ability to build trust because, as outlined by Gambetta, trust inherently depends on agency.

Nissenbaum concludes by emphasising the need to make space for trust to emerge, prompting us to consider when and where to ensure internal security by creating conditions of confidence, and when and where to preserve freedom for human actors in the system. She cites Luhmann and suggests a tradeoff

between reducing complexity through security and enabling people to live with complexity through trust. Security, safety and certainty deliver confidence at the price of limiting space for trust, sociality and complexity.

2.7 Affective Trust

Judith Donath's articulation of "affective trust" in *Unreliable: The Cost of Honesty and the Value of Deception* (forthcoming) considers trust as a means, alongside constraints and sanctions, to achieve more confidence—which Donath defines differently than Luhmann. Donath defines "confidence" as a person's perception of the probability that a claim is true, such as whether a person will act as they say they will. Confidence here is a measure of belief, and the confidence needed to believe a claim varies with perceived risk: if the stakes of being wrong are low, then the threshold of confidence needed to believe a claim is also lower.

Trust is the emotionally rewarding feeling, both for the trustor and the trustee, that another cares about your well-being and thus has aligned their interests with yours; the affective experience is the fuel of prosocial behaviour and fosters a "sense of bonded togetherness" (Donath 5).

By describing trust as a tool to achieve confidence, Donath chooses not to distinguish trust (as the act of transacting *despite* a situation of risk and vulnerability) and confidence (as the act of transacting *because* there is no sense of risk and vulnerability). But she does distinguish affective trust from relations based on past transactions that lower calculated risk, which could be understood as a weak form of confidence-based relations. Most importantly, she clarifies that the choice to increasingly rely on constraints and sanctions to build confidence has consequences for the social fabric, reducing if not replacing the opportunities to establish and sustain trust.

Donath considers hitchhiking and couchsurfing to argue that people are willing to engage in a relationship of trust, without relying on technological guarantees and platform constraints, to the extent that they identify shared values that make them similar in meaningful and relevant ways.

When members of a particular community or institution interact, knowing they share a set of values makes it easy to trust each other. In the cases of hitchhiking or couchsurfing, strangers have not established shared values but likely share a general trust in the benevolence of humanity. This arguably puts hitchhikers and couchsurfers in loosely-tied groups that can facilitate interpersonal trust. While similar people will cooperate without much need for institutional trust-building, people with different values will need greater external support. Donath emphasises the evolving nature of affective trust: we should create more chances for people to explore their relationships and deepen trust with those who were previously more distant from them. Engaging in situations that put people in a position of vulnerability towards a third party contributes to strengthening trust at the interpersonal level, as they enable the trustee to prove to the trustor that they are, indeed, reliable persons—thereby reinforcing the social bond that exists between these two persons.

3 FACTORS IMPACTING TRUST & CONFIDENCE

The following section explores different factors that may contribute to increasing or decreasing trust and confidence. These include restraints on the agency of trustees, the establishment of trust mediators, socio-technical systems, the collective production and assimilation of knowledge, societal pressures, and institutionalisation.

3.1 Restraints on the Agency of Trustees

According to Gambetta's conceptualization of trust as a rational calculation, the decision of whether or not to cooperate based on trust is determined by the constraints, costs and benefits presented by each specific situation. The author distinguishes specifically between two mechanisms to achieve greater cooperation: coercion and lighter constraints, according to their relationship to trust. In line with Nissenbaum and Donath's analyses, Gambetta describes coercion as falling short of being an adequate alternative to trust: expenditures on surveillance, information gathering and enforcement are often too restrictive to the agency of the trustee. These may end up actually reducing trust in the party exercising such coercion, thereby increasing the probability of defection by the coerced.

Other forms of constraints, such as contracts, pre-commitments and promises, may be more compatible with the emergence of trust. These are devices whereby one can impose some restraint on themselves or their trustees in order to reduce the set of alternatives or possibilities for defection. Because agents in these decentralised systems do not know each other and thus cannot trust each other, technological guarantees serve as constraints to ensure others will not defect. Given the reduced agency left to these agents, it is questionable whether the increased confidence provided by blockchain technology could translate into an increased level of trust amongst them. This might depend on whether agents are also involved in the maintenance of the network, such as miners or core developers, who might have a greater sense of agency, leading to trust.

These definitions provide more nuance to our guiding questions. In an environment where trust is ex-ante completely absent, such as digital networks with no embedded identity or reputation systems within them, we have a setting of distrust—which may, as Gambetta points out, generate more distrust. Thus, can we use blockchain technology (or trustless systems) as an artifact for trust to emerge off-chain that can serve as the necessary building blocks to create participatory and cooperative communities?

3.2 Trust Mediators

Balázs Bodó's recent article, "Mediated trust: A theoretical framework to address the trustworthiness of technological trust mediators" (2020), complements thinking on trust by investigating the role of trust mediators. Here, trust is conceptualised as both an interpersonal and institutional relationship, characterised in both cases as an intersubjective relationship. The paper is essentially concerned with how humans trust each other.

The first part of the article argues that globalisation and digitalization have unleashed a crisis of trust. In the realm of globalisation, the crisis is brought forth by the fact that the nation state, as a geographically bounded entity, does not scale trust to the global level. In the realm of digitalization, the crisis is brought forth by the fact that digital technologies are creating a new phenomenon of technologically mediated trust. The need for trust in digitally mediated interactions produced many technical solutions, ranging from various cybersecurity tools and online reputation systems to technological transparency and accountability frameworks, such as open-source code or decentralised architectures.

Approaches to trust-building coalesce into three major categories, each attached with a specific temporality: reputation systems, which mediate the past; blockchain systems, which mediate the present; and recommendation systems, which mediate the future. Reputation systems use signals of trustworthiness that are fine-tuned to the requirements of specific interactions and standardised across highly different local contexts, allowing interpersonal trust to be produced on unprecedented scales and speeds. Such centralised, systematic and algorithmic accounts of reputation transform trust and trustworthiness—social capital—into a commodity that can be quantified, traded, enclosed and sanctioned. Blockchain systems try to minimise the need for trust and produce confidence by hard-coding rules into the system, both at the infrastructure level and in their application (smart contracts). On-chain governance ensures that the behaviour of the system is

predictable. Recommendation systems produce trust from insights often based on inductive statistical analysis or machine learning. Such analysis generates new insights for trustworthy connections and interactions. The article then proceeds by asking how we can establish trust mediators' trustworthiness—and possibly legitimacy Bodó proposes to use the Ability-Benevolence-Integrity (ABI) framework (Mayer et al., 1995) as a possible solution. "Ability" relates to the competence of the trustee to meet expectations of trust. "Benevolence" refers to the trustee's willingness to act beyond their own self-interest and to the benefit of the trustor. "Integrity" signals that the trustor and the trustee share fundamental moral or legal principles, which guide the trustee's actions.

The analysis unveils a problematic scenario for technologically-mediated trust. The current architecture, in many respects, fails to establish its own trustworthiness. Bodó suggests that technological safeguards must be complemented by adequate internal governance structures, and clear external accountability and an embeddedness in local contexts, norms, customs and institutions to reflect local institutional interdependencies of distrust management.

3.3 Socio-Technical Systems

The second chapter of Patrick Sumpf's book, *System Trust: Researching the Architecture of Trust in Systems* (2019), distinguishes system trust from related notions of trust, confidence, and familiarity. In particular, Sumpf breaks from Luhmann who replaced system trust with confidence based on the assumption, which Sumpf rejects, that people do not influence systems and thus systems do not involve risk-taking or decision-making. Sumpf presents an empirical study about the public's relationship with energy provision that reveals traces of specific system trust.

Sumpf recognizes the inherent social component in systems, the ability for human choice, contingencies and risk, and the possibility of having genuine trust in systems. Rather than preferring the term system trust or confidence over the other as Luhmann did, Sumpf disentangles the two notions, seeing them as distinct phenomena.

In fact, Sumpf considers system trust to be more similar to trust in persons, based on what he identifies as the six core components of trust: risk, complexity, control, non-knowledge, suspension of doubt, and—perhaps most central—expectations. Expectations can range from general to specific and be directed at different objects (persons, roles, programs, values), forming an architecture of trust. Within this framework, trust in systems and trust in persons are comparable because they each hold positive expectations for the future, the most significant difference being that system trust is directed at a systemic object (e.g., programs) instead of an individual or group.

Unlike others who find a meaningful distinction between trust as a calculated choice and compulsory trust, Sumpf's socio-technical approach to systems assigns less significance to compulsory trust, in line with our previous authors. To Sumpf, if coercive situations develop and transition over time to a situation of familiarity, they can build trust. For example, it would not be entirely accurate to describe a baby as trusting its mother until it first recognizes her care and becomes willing to put itself in a situation of vulnerability. The baby is vulnerable either way, but the coercive experience builds familiarity, which plants the seed for trust to emerge.

In line with Luhmann, Sumpf's view is that social attributions direct such narratives. There are no objective situations of trust or familiarity: attributions lead our initial actions and those following disappointed trust and thus make a difference for individual and collective trust biographies. Trustors and their environments (trustees, society) attribute (in-)actions differently, leading to either individual responsibility or an experienced compulsion (e.g., agreeing to terms of service on the internet). Emphasising social

constructivist notions, Sumpf empirically grounds system trust in trustors' performance reviews of societal systems such as politics, science, the economy, and socio-technical systems such as energy. The feeling of vulnerability—the sense of the possibility for betrayal—is a normative variable perhaps underestimated regarding system trust in its rather functionalist fashion. Even when systems fail to meet trustors' expectations, some form of agency appears to be required to commit betrayal. Sumpf presents initial plausibility and empirical reconstructions of mediations between a system and its components as trustees. Yet, more research is needed to disentangle the opaque arrangement between a system as such and its parts as possible agents.

Without a related sense of vulnerability, being confident in a system may be more likely than trusting it. Asking how and why people make themselves vulnerable, if they relate to a system or the actors in it, if they form general or specific expectations, and if they act on behalf of more or less knowledge may lead to further clarity.

3.4 Knowledge

Judith Simon's "The entanglement of trust and knowledge on the Web" (2010) considers the underlying trust that we place or withdraw when assessing epistemic claims, especially on the Web where there is so much content that "we have to trust, because we cannot check everything for ourselves" (348).

To an extent, this is the result of our digital systems lacking a notion of truth: the internet is just floating packets and, on the Web, we copy-and-paste without reference or provenance. Next-generation digital infrastructure will, therefore, need a notion of social (and global) truth attestation. Nonetheless, Simon analyzes Wikipedia and other examples to shed light on the communal practises and standards we are able to use today to create social knowledge: "a *social status* that can be ascribed to epistemic content by a community… a *success term* labelling epistemic content that has survived critical scrutiny from multiple agents and satisfies communal standards" (345).

Simon (2010) argues that our trust in Wikipedia is procedural, for its content has survived the continuous enforcement of shared standards among its community of writers and editors. But these processes are complex, and the community relies on various proxies and mechanisms to make them transparent to make epistemic assessments of them. Simon emphasises that these methods must be inspected as well, particularly when epistemic decisions are deferred to illegible algorithms. Our tools should empower people with transparency and the ability to control the amount of trust they place in the content on the web. Simon defines trust broadly, including the possibility for betrayal and disappointment, whereas this report distinguishes trust and betrayal from confidence and disappointment. Given her definition, Simon essentially considers backward-looking trust, or how to trust history and past knowledge, while most literature on trust might be said to consider forward-looking trust, or how to make probabilistic calculations and anticipate future events.

Though both orientations rely on an understanding of the past and involve the calculation of risk, what distinguishes trust is the agency to betray. Knowledge, static and recorded in the past, does not have this agency. A community who accepts and acts on certain knowledge may have agency, but the knowledge itself lacks the freedom to act in the future against the trust one may give to it. Therefore, backward-looking trust in past knowledge is better understood as confidence. Simon rightly argues that knowledge is constructed out of countless layers of trust and confidence all the way down, and that how willing or quickly we accept and assimilate a piece of knowledge, given its underlying architecture of trust and confidence, depends on the stakes and importance of its truth.

If the truth of a claim is important, further research and verification may bolster confidence and lessen

trust—or it may have the opposite effect, as the discovery of more variables, uncertainties and contingencies might heighten the feeling of trust and lower confidence. One should not assume that further analysis and time to make decisions clearly and consistently affect trust and confidence; it depends on the context and consequences. Governance systems could benefit from blending measurements of stake with reputation and wisdom in deciding who is legitimate and worthy of attention, influence and trust.

3.5 Societal Pressures

Chapter 6 of Bruce Schneier's book, *Liars & Outliers* (2012), lists a series of factors—or societal pressures—that can contribute to the establishment of trust and cooperation. To frame his approach, Schneier presents societal dilemmas, also known as coordination games, as competing choices between group interests and individual interests. He writes that there's an alternative, parasitic strategy for any cooperative system: tapeworms in the digestive tract, thieves at a market, or spammers on email. In light of the Prisoner's Dilemma, such parasitic responses are named "defectors." Substantiating his analysis, Schneier argues that such defectors or parasites can only survive if they're not so successful that they completely disrupt the system or the host in which they operate. If tapeworms are too greedy, the host may die, and tapeworms will consequently die with them. Too many thieves in a market will close the market. Too much spam, and no one will read their emails at all. This dynamic creates a fundamental tension between cooperation and defection, between society and individuals. We may not want to pay taxes as individuals, but we're all better off if everyone does. But if everyone cooperates, an individual is better off defecting.

To address this fundamental tension, the author introduces the concept of societal pressures: coordination mechanisms designed to make individual interests align with group interests. This conceptualization is especially game-theoretical. Societal pressures work to modify the structure of payoffs in order to reduce defection: either by increasing the difficulty of defecting, reducing benefits of defecting, limiting the damage caused by defection, or increasing the benefits and lowering the costs of cooperation. His typology of these social pressures includes moral pressures, reputational pressures, institutional pressures and security systems. Moral pressures are "any innate or cultural guidelines," which include informal rewards and penalties and are mostly employed in small group settings. Reputational pressures also entail social or psychological factors but can work in slightly larger groups and use more concrete consequences in order to be effective. Institutional pressures apply to large groups and can be effectively formalised. These are the mechanisms used to solve the "tragedy of the [unmanaged] commons" (Hardin, 1968), though there are limitations to those as well, as social costs circumscribe them.

Security systems are "a weird hybrid" that often fill in the gaps between other social pressures. Capable of physical constraints on human behaviour, they are pervasive in every realm of life as tools to augment any other societal pressures. His description of these as pathways to trust blurs this report's distinction between "trust" and "confidence." But understanding the social embeddedness of trust, its underlying affective experience, and the ability to control populations reveals much more of its complexity. As described by Donath, substituting the affective experience of trust with confidence, achieved either through external constraints or technological means, eliminates all the benefits of trusting relationships between people. Couchsurfing requires people to invest lots of time into developing personal relationships and trusting fellow Couchsurfers. There is no opportunity to establish trust on a platform like Airbnb because there are constraints imposed from above and below (rules and security). Then again, embracing trust alone has a different set of problems associated with it. For instance, due to homophily, we find it much easier to trust people who are similar to us. Building base layers of confidence may encourage more trusting interactions and cooperation across differences.

3.6 Institutionalisation

In "Trust, Confidence and Voluntary Organisations: Between Values and Institutions" (1999), Tonkiss and Passey claim that trust relations are ethical relations based on shared values and conditioned not by an external framework of controls but by agency and voluntarism. These are distinct from confidence relations, which are secured by contract or other regulatory forms. The authors put forward a different understanding of system trust to account for the type of system involved, explaining that secondary organisations, which range in scale from family to state, work to generalise trust. But the authors find a "potentially contradictory relationship" between their institutional form as enterprises and their underlying values (Tonkiss, Passey 272). Their core argument is that a "changing operating environment — marked by greater regulation, competition, professional fundraising, corporate sponsorship and a 'contract culture' means that voluntary organisations are increasingly governed by mechanisms designed to develop and maintain confidence" (Tonkiss, Passey 272). The consequence of mediating voluntary relations through confidence-based measures is the erosion of trust, which the authors consider to be the basis for voluntary organisations and their caring functions. An organisation built on trust reinforces such trust and can even generate further and deeper trust, whereas a more institutional arrangement relies on and develops more confidence.

In place of trust relations built on shared evaluations of the social good, are hardened and formalised confidence relations built on procedural and target-driven contracts. This creates tension between "doing good" on social causes with acts of care and "doing well" as a competitive organisation with more financial and administrative work. The consequence is that "philanthropic discourses of need which are geared towards generally desired social outcomes are redefined in the language of precise technical "outputs" (Tonkiss, Passey 268).

Vicious cycles of institutionalisation and bureaucratization—regulating, conditioning, and contracting all interactions—clearly crowd out space for trust. Proper institutions, however, find ways to harness trust and confidence together as reinforcing complements.

Most interactions already sit somewhere along the continuum between absolute trust and watertight, contractual confidence. In fact, an institutional shift does not necessarily weaken trust relations. Through guaranteed execution and transparency, providing confidence enables interactions that would not otherwise be possible, which creates new opportunities to build trust.

These are fractal and concentric systems of trust and confidence. Workplace relations, for instance, can be strongly trust-based, built upon bureaucratic but also social forms of interactions. Or consider childcare relations: an immediate family member will care for a child, for free, based on a relationship of fundamental and unconditional trust. Childcare provided through intermediate associations of state forms is increasingly a relationship of confidence. Still, a few would keep a caretaker who at some level they did not trust—such trust might simply be an effect of the initial confidence relations allowing establishing a trust relationship.

4 TRUST & CONFIDENCE IN BLOCKCHAINS

As some of us have argued elsewhere (De Filippi, Mannan & Reijers, 2020), blockchain technology creates confidence about its operation based on cryptographic primitives and game-theoretical incentives. While these properties strengthen the security of public, permissionless blockchains by minimising the risk of tampering with transactions, as Nissenbaum observes in her paper (106-108), security does not adequately address trust' functions, which include diminishing complexity and coping with risk. "Trust does not flourish in a perfectly secure environment" in Nissenbaum's view, as it lacks uncertainty, the freedom to choose, and the possibility of proving a trustee's trustworthiness, all of which are inherent in trust

relationships (Nissenbaum 123).

While negating the possibility of there being breaches of trust (e.g. by banks) and reducing risk may be the very reason why some people are drawn to use blockchains, Nissenbaum argues that security provided by a blockchain network only protects a socio-technical system from "outsiders", but does little to guard against "insiders" (e.g., miners, core developers) who may act in self-serving and socially harmful ways (Nissenbaum 125). In short, risk and uncertainty remain, particularly from within (concerning off-chain governance). This is especially true of public and permissionless blockchain-based networks where it may not be easy to distinguish between outsiders and insiders. Thus, to guarantee the proper governance and functioning of a blockchain-based network, it is useful to question if we want to build more trust amongst the network nodes and the network participants, or if we want to increase the "trustless" character of the network by increasing confidence at the off-chain governance layer.

As Gambetta notes, there are situations in which it is rational to trust trust and others in which it is not. Gambetta contends that one has to be open to the possibility that others will be trustworthy, as this openness to trust may in itself motivate trustees to behave in a trustworthy manner. Still, ultimately the decision to repose trust should be premised on evidence-based, rational calculation. The decision to trust should only be made if this calculation finds that the trustee's interest encapsulates the trustor's interest (Hardin 11). According to Hardin, the trustee will only have an interest in being trustworthy if they wish to continue a relationship with the trustor (Hardin 1). In a Prisoners' Dilemma game, for example, such a rational approach would suggest that there are no bases for trust and it would be better off for both the potential trustor and trustee to not cooperate.

Ordinarily, as Schneier argues, security measures, laws, social norms and other societal pressures can allow us to trust and cooperate—thereby changing the payoff structure of Prisoners' Dilemma games—because they provide the institutional and informal constraints that inhibit defection in different ways (Schneier 64). These societal pressures are not present in the same manner in blockchain-based networks. There are options for building trust in off-chain governance, but each option has limitations due to the decentralised and pseudonymous nature of public, permissionless blockchains.

First, institutional pressures are challenging to implement in a decentralised system with no possibility for coercion. On one hand, the global dispersion of a blockchain network makes it challenging for territorial legal systems to exercise institutional pressure on the entirety of the network. On the other hand, any institutional rules established by a particular community or organisation will lack a mechanism to enforce its applications by all network nodes. The question is whether it is possible, or even desirable, to adopt such an approach, as that would be incompatible with a decentralised governance structure.

Second, reputational pressures also have limitations as a trust-enhancing mechanism for large-scale, public and permissionless blockchain networks, which are often characterised by a strong degree of pseudonymity. Reputational pressures work best in small or medium-sized groups, with people engaging in repeated interactions (Schneier 71). Pettit's related conceptualization of trust as reliance—the desire to be seen by others as trustworthy—is similarly unlikely to emerge at the on-chain level, due to the consensus mechanisms, pseudonymous nature of interactions, and strongly codified and predetermined environment affording little agency to actors. Nevertheless, this conceptualization may be relevant at the off-chain governance level. When the community needs to deliberate over the algorithmic rules that will govern the blockchain system at hand, one can rely on the fact that others want to be perceived as behaving in a trustworthy fashion.

And third, shared values help build trust, but in a pseudonymous environment it might be difficult to assess

whether a potential trustee shares the same moral and social values as the trustor.

In on-chain governance, there are security measures and encapsulated group interest in the blockchain-based network being valuable (however defined), but moral, reputational and institutional pressures (Schneier 69-70) are either weak or absent. In off-chain governance, there are strong moral, reputational and institutional pressures but weak security measures. These societal pressures do not provide sufficient confidence in the off-chain governance system.

Hence, off-chain governance currently relies more on trust than confidence. Nguyen argues that trust as an unquestioning attitude and the integration of the trusted can be useful, as it creates cognitive efficiency, and may even be a necessity as individuals with finite cognitive resources in a complex world. But it also creates vulnerability, begging the question of the appropriate criteria for suspended deliberation. Perhaps Nguyen's core question moving forward is how to secure trust without the expertise and cognitive resources to secure it perfectly.

On that point, it might be helpful to consider whether alternative mechanisms of confidence-building could be devised at the off-chain governance level in order to increase the confidence in its operations, rather than relying only and exclusively on trust. Any legitimate, polycentric system of governance will have to find a proper combination of confidence and trust so that risk diminishes and the blockchain operates securely, while still leaving room for agency and freedom for action. Donath's notion of affective trust has value in and of itself, and we should aim to cultivate it.

We can initially rely on constraints and sanctions to create low-risk situations and more confidence in Luhmann's sense, and—over time—introduce more agency, risk, and vulnerability into the shared spaces so as to foster trust. Hence confidence, through constraints and sanctions, can actually help strengthen the social fabric—supplementing and bootstrapping trust, but not replacing it. Ultimately, the limited forms of coercion available in decentralised systems, particularly in blockchain-based networks, make it all the more important for the on-chain and off-chain governance of the system to be perceived as being legitimate.

In "Public Confidence in Policing: A Neo-Durkheimian Perspective" (2007), Jackson and Sunshine investigate the social and psychological forces that determine community levels of trust, legitimacy and consent toward police. In line with this report's conclusion, they frame trust and confidence in a broader context as sources of legitimacy. To the authors, legitimacy is about the dyadic relationship between an institution or other power-holder and the people over whom it makes moral and normative claims to power; there are public expectations that define what the holder of power should do in order to be seen by the audience as moral, just and appropriate. In the context of policing, the authors contribute illuminating empirics to the concepts of this report, having conducted their study in rural England. Their core claim is that legitimacy rests on fair procedures and outcomes "or on the belief that the criminal justice system defends and expresses foundational values" (218). They investigated attitudes toward procedural justice and the values that the police represent, finding that public acceptance and, thus, the legitimacy of police derives not from low fear of risk and crime but rather from feelings about whether the police share the community's values and morals. If the police exemplify those values and morals and ensure fairness and human dignity in their conduct, they garner legitimacy. The police represent the institution of policing, which people associate with moral order and social stability. As such, beyond their effectiveness or their procedural justice, the police are legitimate if the neighbourhood is thought to be orderly, cohesive, and trusting—a place where people act for the common good. This means that in a self-policing area with collective efficacy, the police—being the formal institution of policing—earn legitimacy for the success of informal policing. If the legitimacy of police depends less on brute force and more on the authority they command through public trust, then a loss in legitimacy may actually come from the deterioration of shared values more than of public safety. An institution that loses legitimacy often falls back on coercion to

preserve itself, and in a normal policing environment, there is no ability to exit without high cost: you may be stuck with the coercive, illegitimate institution. But the effects on trust, confidence and legitimacy may be different in a purely voluntary, blockchain-based system with a low cost to exit. If trust is the voluntary delegation of power to someone who is thought to act with aligned interests and values, then it can be a source of legitimacy. If we trust a person that has power, we might consider their actions and decisions to be more legitimate. But while we voluntarily give trust to someone, the authors' conceptualization of legitimacy primarily concerns consent.

This brings into question how levels of trust and confidence inform the perceived degree of legitimacy: can a power dynamic be legitimate if we did not willingly choose to put ourselves in a position of vulnerability toward the actor with power over us? Low exit costs would help ensure accountability and possible intervention if legitimacy comes into question. In fact, in a voluntary system with easy exit and no coercive authority, legitimacy is a necessary force for the system's sustainability. But when a system, like policing, becomes essential to the social fabric—increasing exit costs—then the ability to include, participate and express voice becomes essential to building persistent yet evolving legitimacy.

5 CONCLUSION

This report analyses the interrelationship between trust and confidence in the context of blockchain technology. It presents the distinction between these two terms, summarises divergent conceptualizations of trust, and highlights a series of factors that can increase or decrease trust and confidence. Maintaining that blockchain technology is not a trustless technology but rather a *confidence machine*, the aim is to foster discussions on how different levels of trust and confidence, both on-chain and off-chain, can influence the perceived degree of legitimacy of blockchain-based networks.

Due to the decentralised and pseudonymous nature of permissionless blockchains, legitimate polycentric systems of governance need to find a proper combination of confidence and trust so that risk diminishes and the blockchain operates securely, while still leaving room for agency and freedom for action. Further research is needed to understand what makes decentralised governance systems legitimate (Buterin 2021), a topic that will receive greater attention in a forthcoming paper: How do trust and confidence play a role in establishing legitimacy in the governance of a polycentric system?

6 REFERENCES

Bodó, B. (2020). Mediated trust: A theoretical framework to address the trustworthiness of technological trust mediators. *New Media & Society, 23*(9), 2668–2690. https://doi.org/10.1177/1461444820939922

Buterin, V. (2021). *The Most Important Scarce Resource is Legitimacy*. Vitalik.ca. https://vitalik.ca/general/2021/03/23/legitimacy.html

Coeckelbergh, M. (2012). Can we trust robots?. *Ethics And Information Technology, 14*(1), 53-60. https://doi.org/10.1007/s10676-011-9279-1

De Filippi, P. & Mcmullen, G. (2018). *Governance of blockchain systems: Governance of and by Distributed Infrastructure* (Research Report). Blockchain Research Institute and COALA. Hal02046787f

De Filippi, P., Mannan, M., & Reijers, W. (2020). Blockchain as a confidence machine: The problem of trust & challenges of governance. *Technology In Society, 62*, 101284. https://doi.org/10.1016/j.techsoc.2020.101284.

Donath, J. (2021). (forthcoming), *Unreliable: The Cost of Honesty and the Value of Deception*.

Gambetta, D. (2000). Can We Trust Trust. In D. Gambetta (Ed.) *Trust: Making and Breaking Cooperative Relations* (pp. 213-237). Blackwell.

Hardin, Garrett (1968), The Tragedy of the Commons. *Science, 162*, 1243-1248.

Hardin, R. (2002). *Trust and Trustworthiness*. Russell Sage Foundation.

http://www.jstor.org/stable/10.7758/9781610442718

Jackson, J., & Sunshine, J. (2006). Public Confidence in Policing. *The British Journal Of Criminology, 47*(2), 214-233. https://doi.org/10.1093/bjc/azl031

Luhmann, N. (2000), Familiarity, Confidence, Trust: Problems and Alternatives. In D. Gambetta (Ed.). *Trust: Making and Breaking Cooperative Relations* (pp. 94-107). Blackwell.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review, 20*(3), 709–734. https://doi.org/10.2307/258792

Nguyen, C. Thi (forthcoming), Trust as an unquestioning attitude, Oxford Studies in Epistemology. https://philarchive.org/archive/NGUTAA

Nissenbaum, H. (2001). Securing Trust Online: Wisdom or Oxymoron? *Boston University Law Review, 81* (3), 635-664. https://ssrn.com/abstract=2573181

Pettit, P. (1995). The Cunning of Trust. *Philosophy & Public Affairs, 24* (3), 202–225. http://www.jstor.org/stable/2961900

Schneier, B. (2012). *Liars and outliers: Enabling the trust that society needs to thrive*. Wiley.

Simon, J. (2010). The entanglement of trust and knowledge on the Web. *Ethics And Information Technology, 12* (4), 343-355. https://doi.org/10.1007/s10676-010-9243-5

Sumpf, P.(2019). *System Trust: Researching the Architecture of Trust in Systems*. Springer VS.

Tonkiss, F., & Passey, A. (1999). Trust, Confidence and Voluntary Organisations: Between Values and Institutions. *Sociology, 33* (2), 257-274. https://doi.org/10.1177/s0038038599000164