Philosophy & Technology (2021) 34:573–587
https://doi.org/10.1007/s13347-020-00410-x

RESEARCH ARTICLE

How Implicit Assumptions on the Nature of Trust Shape the Understanding of the Blockchain Technology

Mattis Jacobs[1]

Received: 16 December 2019 / Accepted: 14 June 2020/ Published online: 29 June 2020

Abstract
The role that trust plays in blockchain-based systems is understood and portrayed in various manners. The blockchain technology is said to enable and establish trust as well as to redirect it, to substitute for it, and to make it obsolete. Furthermore, there is disagreement on whom or what users have to trust when using the blockchain technology: (only) code, math, algorithms, and machines, or still (also) human actors. This paper hypothesizes that the divergences of the depictions largely rest on implicitly adhering to different accounts of trust. Thus, the goal of this paper is to outline how the current lack of a shared understanding of the term "trust" leads to diverging interpretations of the blockchain technology's core features. Furthermore, it shows how this lack of common understanding obstructs scholars from referring to one another meaningfully in the discourse on blockchain technology. To do so, this paper outlines the most prominent depictions of the setup of relevant trust relationships within blockchain-based systems and traces their roots to different underlying assumptions on the nature of trust.

Keywords Trust · Trustworthiness · Blockchain · Bitcoin, distributed ledger technology

1 Introduction

On November 1, 2008, an author or a group of authors under the pseudonym "Satoshi Nakamoto" published a whitepaper titled *Bitcoin: A Peer-to-Peer Electronic Cash System* (Nakamoto 2008) on a cypherpunk mailing list. It outlined a novel approach for enabling cryptocurrencies, apparently free of centralized

✉ Mattis Jacobs
jacobs@informatik.uni–hamburg.de

[1] Department of Informatics, Universität Hamburg, Vogt-Kölln-Straße 30, 22527 Hamburg, Germany

authority and without roots in incumbent institutions. The implementation of Bitcoin followed in 2009. In the coming years, it gained significant attention as a proof of concept for "the next step in the evolution of money" (Maurer, Nelms, & Swartz 2013, p. 273). However, the underlying blockchain[1,2] technology quickly outgrew the application area of cryptocurrencies. Subsequent blockchain-based systems incorporate Turing-complete virtual machines instead of supporting only very limited scripting languages. Thereby, they do not only allow the digital transfer value without relying on the third-party intermediaries (cf. Swan & Filippi 2017) but also enable so-called self-executing smart contracts, decentralized applications (DApps), decentralized autonomous organizations (DAOs), and several other novel phenomena and organizational structures (Buterin 2014).

Thus, as *The Economist* (2015) keenly observed regarding the blockchain technology, "[t]he real innovation is not the digital coins." Instead, the article identifies the unusual role of trust in blockchain-based systems as the outstanding element across all application areas. However, while many scholars share this view (Antonopoulos 2017; Beck et al. 2016; Filippi 2017; Hawlitschek et al. 2018; Mallard et al. 2014; Werbach 2018), the role that trust plays in these systems is understood and portrayed in various manners. The blockchain technology is said to enable (Underwood 2016, p. 16) and establish (Krishna 2015) trust as well as to redirect it (Werbach 2018, p. 30), to substitute for it (Freeman et al. 2020, p. 69), and to make it obsolete (Nakamoto 2008, p.8). Furthermore, there is disagreement on whom or what users have to trust when using the blockchain technology: (only) code, math, algorithms, and machines (Maurer et al. 2013; Nakamoto 2008), or still (also) human actors (Botsman 2017; Walch 2019; Werbach 2018). While some depictions of the role of trust in blockchain-based systems prove to be dominant in the discourse, no agreeable and comprehensive one has asserted itself to this day.

This paper hypothesizes that the divergences of the depictions largely rest on implicitly adhering to different accounts of trust. As Hardin (2002, pp. 87–88) notes more generally regarding discourses around trust, if they "are to be understood, [participants in the discourse] must specify more narrowly how [they] mean to use the term." Thus, the goal of this paper is to outline how the current lack of a shared understanding of the term "trust" leads to diverging interpretations of the blockchain technology's core features. Furthermore, it shows how this lack of common understanding obstructs scholars from referring to one another meaningfully in the discourse on blockchain technology. To do so, this paper outlines the most prominent depictions of the setup of relevant trust relationships within blockchain-based systems and traces their roots to different underlying assumptions on the nature of trust.

[1] If not specified differently, the term "blockchain" in this paper only refers to open, permissionless systems. Furthermore, only direct interactions with blockchain-based systems are taken into account. Because second layer applications, i.e., applications building on top of these systems, do not necessarily share all the relevant features with the systems they are based on, they are not considered in this paper.

[2] In most academic literature, the term "blockchain" increasingly gets supplanted by broader terms like "distributed ledger technology" or "append-only databases." However, these terms include also systems with similar characteristics but different operating principles. Since the trust issues discussed in this paper largely depend on the operating principle, the line of reasoning and the results cannot necessarily be transferred to those systems. Therefore, the term "blockchain" is still applied in this paper.

2 Depictions of the Role of Trust in Blockchain-Based Systems

The Bitcoin Whitepaper does not merely outline the technical foundations of the blockchain technology but also provides an interpretation of the role of trust in it. Nakamoto characterizes the (Bitcoin-)blockchain as trust free, i.e., he suggests users can use it "without relying on trust" (Nakamoto 2008, p. 8). However, as the discourse matures, the notion of a trust-free technology is often used in a narrower sense and refers only to one of two things. On the one hand, the characterization is used to suggest that the necessity to either trust transactional counterparties or intermediaries vanishes when using the blockchain technology. This necessity usually exists when transferring assets with traditional payment processors. For instance, Swan (2015, xii) follows this line of reasoning and describes trust-free transactions as "at its most basic level, intermediary-free transactions."

On the other hand, there is a temporal dimension. The alleged trust-free nature of the blockchain technology also manifests in its capacity to determine future action. Without the application of blockchain technology, users have to trust external entities to perform certain actions in the future, as, e.g., enforcing contracts or controlling the money supply in a desired way. In contrast, the blockchain technology promises to predetermine such actions. It enables users to create self-enforcing contracts and use a currency that is "produced at a predictable rate, with a maximum number [of tokens] preestablished" (Christopher 2016, p. 172). Therefore, DuPont and Maurer (2015, p. 9) argue that the blockchain technology "seeks to put boundaries around uncertainty"—a sine qua non of trust—"to the point of snuffing it out." However, this line of reasoning refers to specific features of the blockchain technology. It does not allege that users of blockchain-based systems do not encounter uncertainties or trust issues at all.

In contrast, the depiction of blockchain as a technology based on trust in code, math, or algorithms suggests that blockchain-based systems do not eliminate the need to trust at all. Instead, it suggests that there is a shift concerning whom—or what—users have to trust when using blockchain-based systems in comparison with whom or what they have to trust when transacting and interacting by other means. Just like advocates of the depiction of blockchain as a trust-free technology, advocates of this depiction also assume that trust becomes obsolete in some areas. For instance, Maurer et al. (2013, p. 264) suggest that for "Bitcoin to work, one does not have to trust Nakamoto, a bank, or any other person or institution." However, in contrast to the depiction of blockchain as a trust-free technology, this depiction of blockchain technology does not end with the determination of where relationships based on trust become obsolete. Instead, it outlines what they are replaced with. Here, Maurer et al. (2013, p. 264) note that instead of trusting intermediaries, one "must simply trust the code or, more precisely, the cryptographic algorithm."

The third depiction assumes that the role of trust in blockchain-based systems differs from trust in other setups in that it is placed in networks of actors instead of individual actors (Werbach 2017, p. 501). For this co-founder of LinkedIn Reid Hoffman coined the term "trustless trust." He suggests that the setup of trust relationships within blockchain-based systems comprises relationships of a novel nature in which no individual can be identified as the sole trustee (Hoffman 2014). Another commonly used term to describe this phenomenon is "distributed trust." This insight constitutes the basis for Werbach's seminal book *The Blockchain and the New Architecture of Trust*

(2018), potentially the most comprehensive work on the issue so far. Based on the understanding of the blockchain technology as a facilitator of distributed trust, Werbach sheds light on the differences between the role of trust in blockchain-based systems and the role of trust in other predominant setups in society. Here, he references peer-to-peer trust, i.e., interpersonal trust between transacting individuals; "Leviathan" trust, i.e., trust that is established by a "powerful central authority operates largely in the background to prevent others from imposing their will through force or trickery"; and intermediated trust, i.e., trust that is established through the internal rules "and the reputation of the intermediaries" (Werbach 2018, p. 27).

3 Diverging Assumptions on the Nature of Trust

The term "trust" is often insufficiently defined in publications discussing the role of trust in blockchain-based systems. However, trust is a multi-faceted term, and its meaning is not always evident in a given publication. This especially holds true for interdisciplinary contexts. As shown in the following, one reason for the emergence of diverging views on the role of trust in blockchain-based systems is the adherence to different accounts of trust. This section introduces shared as well as conflicting assumptions among various accounts of trust in order to illustrate how using the term in one way or another changes the understanding of the role of trust in blockchain-based systems.

An assumption at the core of most accounts of trust is its basic structure as a three-place predicate of the form A trusts B to/with X, where *A* and *B* constitute actors—the trustor (or truster) and the trustee—and *X* and action, testimony, or "valued thing" (Baier 1986, p. 236; cf. Hardin 2002, p. 9). Furthermore, trust is a reductive term in the sense that it "is not a primitive, something that we just know by inspection, as the color blue might be a primitive […]. Rather, it is reducible to other things that go into determining trust" (Hardin 2002, p. 57). What these "other things" or components of trust are, however, scholars disagree. Most accounts of trust assume that trust requires the trustor to "1) be vulnerable to others […]; 2) think well of others, at least in certain domains; and 3) be optimistic that they are, or at least will be, competent in certain respects" (McLeod 2006). *Less demanding* accounts of trust stop here. For instance, Gambetta (1988, p. 217) defines trust in this line of thinking as "a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action […]." Trust, according to less demanding accounts, is a matter of *rational expectations* and is often equated with mere judgments of reliability.[3]

However, another school of thought claims that trust and mere judgments of reliability are disparate concepts. It suggests that trust is a *more demanding* concept. Thus, according to richer, i.e., more demanding accounts of trust, for an attitude to qualify as trust, it needs to meet more requirements. In other words, more demanding

[3] In fact, trust according to these accounts is most often not compared with judgments of reliability but with reliance. However, as Nickel (2013, p. 224) observes, "this is not a suitable comparison. Reliance is way of acting, whereas trust is an attitude." This paper follows Nickel's reasoning that the appropriate attitudinal counterpart for trust is a "judgment of reliability."

accounts of trust define trust as a judgment of reliability + x. However, what the "magic ingredient which distinguishes (dis)trust from mere (non-)reliance" (Hawley 2017, p. 231) is even scholars who call for the distinction disagree. Candidates are, e.g., certain motivations such as goodwill towards the trustor (Baier 1986), encapsulated interest, i.e., the idea that trusting requires "a commitment to acting at least partly in the interests of the truster because they are the interests of the truster" (Hardin 2002, p. 57), the appropriateness of specific reactive attitudes such as feeling betrayed in case of misplaced trust (Baier 1986), or the necessity of the trustor to see "a trustee as morally obligated, committed, or accountable in appropriate ways" (Hawley 2017, p. 231; see also Simon 2013).

The requirements of both more and less demanding accounts of trust are mostly discussed with regard to interpersonal settings, i.e., for trust relationships in which the trustor and the trustee are individual human actors. Accounts of trust that consider other entities, potential trustees are usually defined by outlining their divergence from interpersonal trust as the initial and basic concept. Examples of such accounts are, e.g., accounts of trust in groups and organizations (Hawley 2017), accounts of trust in governments (Hardin 2002), or accounts of trust in technological systems (Nickel 2013). Less demanding accounts of trust based on rational expectations are more readily applicable to non-interpersonal settings.

However, in blockchain-based systems, even concerning the relationships among human actors, it is not clear whether a demanding account of interpersonal trust can be applied. This is due to specific technical features of open and permissionless blockchain-based systems in which users are represented through digital keys. The resulting pseudonymity has the effect that users do not necessarily know with whom they are engaging. This eliminates the possibility to evaluate "contextual features" (Werbach 2018, p. 29) of contractual counterparts and validators, i.e., miners. This non-identification goes along with an open validation process that does not depend on excluding actors that do not ex ante prove to be trustworthy (Antonopoulos 2014).

Thus, if the relationship of the two contracting parties or between users and miners is characterized as a relationship based on trust, a less demanding account of trust that does not require a "grounding in specific prior or subsequent relationships with those others" (Hardin 2002, p. 60) needs to be applied. Hardin (2002, p. 62), who advocates a more demanding account of trust, rejects the idea of such generalized trustees. He alleges that the respective propositions do not really claim "that one trusts those others, but only that one has relatively optimistic expectations of being able to build successful relationships with certain, perhaps numerous, others […]."

Taking one or the other side in the debate on what qualifies as an instance of trust is decisive for how relationships among different human actors in blockchain-based systems are to be characterized. If a more demanding account of interpersonal trust is applied, the relationships between transacting parties as well as between users and miners cannot be considered to be relationships based on trust. Instead, the respective stances could be characterized as mere judgments of reliability. Taking this perspective, the characterization of blockchain-based systems as trust-free appears to be much more tenable, without even touching the issue of whether and, if so, where exactly vulnerabilities and uncertainties vanish. If dealing with the many vulnerabilities and uncertainties existing in blockchain-based systems requires only mere judgments of reliability, then this points to the validity of the argument that trust plays a lesser role in

blockchain-based systems compared with other setups. Yet, if blockchain-based systems are indeed entirely trust-free remains to be shown. This issue is addressed again later on.

Similar questions arise regarding whether or not regulatory bodies (i.e., governmental institutions), core developers (i.e., more or less institutionalized groups), the distributed network of miners, and newly established intermediaries like cryptocurrency exchanges (i.e., organizations) qualify as potential trustees. While large proportions of the literature on trust focus on interpersonal trust, some scholars also introduce accounts of trust that are non-interpersonal, "including 'institutional trust' (i.e., trust in institutions), trust in government" (McLeod 2006) as well as in technological systems (Nickel 2013), groups and organizations (Hawley 2017), and many more. Therefore, the accounts of trust in the above-mentioned entities are worth assessing, even though requirements of more demanding accounts stemming from interpersonal settings do not prima facie appear to be applicable.

Less demanding accounts of trust based on rational expectations that equate trust with judgments of reliability are generally open to being applied in contexts where the trustee is not a human actor but, e.g., an organization such as a cryptocurrency exchange. According to such less demanding accounts of trust, it could be sufficient to "be confident that the design of the roles and their related incentives will get role holders to do what they must do if the organization is to fulfill our trust." Still, a reasonable judgment would require having a clear understanding of the structures of and the roles within an institution "to be confident of the incentives or other motivations that foster trustworthiness among role holders" (Hardin 2002, p. 156). If, on the contrary, a more demanding account is applied, it is questionable whether such entities qualify to constitute a trustee. For instance, Hardin's account, which focuses on encapsulated interest, is incongruous with such entities as trustees. According to his account, the trustor would be required to "know that the agents or the institution act on [the trustors] behalf because they wish to maintain their relationships with" them (Hardin 2002, p. 156). For larger institutions, he argues that this is "generally not possible."

However, the matter is contested. Hawley (2017) makes the counterargument that the distinction between trust and mere judgments of reliability matter at the individual level but less so at the level of collective entities. Collective entities like institutions—in contrast to individuals—can have an obligation to be reliable. She claims that there "is no general obligation upon individuals to be reliable, which is why we need the language of trustworthiness to highlight those particular respects in which individuals are obliged to be reliable. Nevertheless, we can require of our institutions that they be reliable in the respects that matter to us […]." Therefore, negative reactive attitudes, such as feeling betrayed, can be reasonably applied in the case that certain institutions prove to be unreliable. This, according to Baier (1986) and others, is only appropriate in contexts where genuine trust is required. Hawley therefore argues for abandoning the distinction between mere judgments of reliability and trust on the level of collective entities. The abandoning of the distinction thus opens the door for a morally laden account of trust regarding organizations and institutions. Accordingly, whether or not groups, organizations, and institutions qualify as trustees varies even among different more demanding accounts of trust.

The answer to the question of whether or not technological artifacts can constitute trustees appears to be more clear-cut. If a more demanding account of trust that exceeds judgments of reliability based on rational expectations is applied, claims of trust in

technology can only be understood as trust in the human actors behind the technology (Holton 1994, p. 66; Jones 1996, p. 14; Nickel 2013, p. 224). Technological artifacts themselves do not have intentions or motivations concerning the actors who assess their reliability. Moreover, actors who assess their reliability usually do not assume this. According to more demanding accounts of trust, technological artifacts are "paradigmatic examples of things about which we make judgments of reliability rather than things we can genuinely trust" (Nickel 2013, p. 224). Even though less demanding accounts of trust which do not differentiate between trust and mere judgments of reliability mostly stem from the idea of interpersonal trust, they are generally also open for technological artifacts as trustees. If trust is equated with judgments of reliability, technological artifacts are trusted if an actor assesses their reliability positively.

Thus, assumptions on the nature of trust have significant implications on how the role of trust in blockchain-based systems can be conceptualized. Some depictions of "the trust revolution of the blockchain and distributed ledger technology" (Werbach 2018, p. 30) are evidently only compatible with a specific account of trust, even if the adherence to the account is not made explicit. According to more demanding accounts of trust, fewer stances are considered genuine trust, whereas, according to less demanding accounts of trust, more stances are considered genuine trust. Describing the shift as one from human actors (groups and institutions or individuals) as trustees to "algorithms that govern users' interactions" (Hawlitschek et al. 2018, p. 57), a "cryptographic algorithm" (Maurer et al. 2013, p. 264), "the instrumental operation of mining" (Velasco 2017, p. 722), "an open source code" (Atzori 2015, p. 7), or "collectives of machines" (Werbach 2018, p. 30) are incongruous with more demanding accounts of trust that presume specific moral or attitudinal components a prerequisite. Trust in the mining community and other collective entities whose members are generally assumed to act exclusively based on self-interest (Werbach 2017, p. 504) also appear to be incompatible with at least some of the more demanding accounts of trust.

4 On the Boundaries of Blockchain-Based Systems

The conceptual matters regarding the nature of trust are not a standalone issue. Adhering to different accounts of trust in the assessment of the role of trust in blockchain-based systems requires focusing on different relationships among actors. For instance, basic assumptions of scholarly disciplines can entail specific conceptions of trust. As shown in the previous section, these conceptions determine whether or not specific components of a socio-technical system are considered potential trustees. If components of a specific type, e.g., cryptographic algorithms, are considered potential trustees, a comprehensive analysis of trust relationships among entities in blockchain-based systems requires the incorporation of components of this type in the respective investigation. This section outlines the varying approaches to including and excluding different actors and entities in the assessment of the role of trust in blockchain-based systems.

The most crucial distinction regarding this issue is between, on the one hand, inquiries that look at blockchains as either closed ecosystems or technical models with strictly defined boundaries and, on the other hand, inquiries that consider currently implemented solutions and the broader environment that these solutions are embedded

in. The latter inquiries take into account many more actors that users can potentially be vulnerable to, e.g., cryptocurrency exchanges, regulators, developers, and the human actors behind the cryptographic keys that represent users. Furthermore, they also often include interactions that exceed the boundaries of the narrower technical system and thus do not entirely fall in the purview of the technology's security features. Instances of exchanges that exceed the technical boundaries are, e.g., exchanges of on-chain assets (e.g., cryptocurrency tokens) for off-chain assets (e.g., Fiat money). Thus, the stricter the focus is on the narrower technical system, the fewer vulnerabilities and uncertainties are taken into account. Accordingly, the more sensible depictions of the blockchain technology as being trust-free or only based on trust in technological components appear.

More technical literature highlighting the innovative nature of the blockchain technology often has this rather narrow scope and ignores actors such as cryptocurrency exchanges that raise severe trust issues. Examples of such elaborations can be found in the Bitcoin Whitepaper. While Nakamoto's outlining of his motivation to develop Bitcoin sometimes transcends the boundaries of technical modeling and takes societal aspects into account, the statements regarding the Bitcoin blockchain as "a system for electronic transactions without relying on trust" remain relatively strictly within these boundaries (Nakamoto 2008, p. 8). The same can be said about Antonopoulos (2014) and his depiction of the blockchain technology as being based on trust by computation. His elaborations also stay within the boundaries of technical modeling, and the discussion of the role of trust omits taking the broader environment of blockchain-based systems into account.

However, scholars such as Botsman and Werbach broaden the scope and take the fringes of blockchain-based systems and the (socio-)technical layers underlying them into account. While acknowledging trust minimizing features within the narrower technical system, they corroborate that humans are still very much in the loop in blockchain-based systems and they can exert power individually, i.e., in a non-distributed manner. Botsman (2017) here itemizes "programmers, […] entrepreneurs and experts who establish and maintain the cryptographic protocols." The list can be complemented by regulators, cryptocurrency exchanges, providers of the underlying internet infrastructure, and many more (see de Filippi and Wright 2018). Since these actors do not operate according to the blockchain protocol but exert power by other means, potential trust towards them is not based on the game theoretical assumptions underlying the protocol's incentive system.

Thus, if the broader environment of blockchain-based systems is taken into account, issues neglected in a characterization that considers only the actors within the narrower technical system become visible. Regularly, for instance, cryptocurrency exchanges and their users fall as victim to hacks, attacks, and frauds (Chohan 2018). Since they are for most users next to impossible to circumvent and users are highly vulnerable through and towards them, uncertainties (which one might argue require trust to overcome) to some degree thwart the steps taken within the narrower system to move away from the need to trust intermediaries. Furthermore, if (core-)developers and their abilities to assert changes to the system's protocol are also taken into account, it is necessary to consider the system itself an ever-evolving rather than a static entity. Therefore, especially in the case of transactions with a longer settlement duration, uncertainties resulting from updates of the system's protocol play a prominent role. The maintenance of the features of the technical system themselves,

including the ones responsible for the alleged trust-minimizing features of the narrower system, are dependent on human actors, which, as Walch (2019) argues, users have to trust in turn. Thus, if the broader environment of the technical systems is taken into account, it is accordingly possible to identify trust relationships between users and other (human) actors that can be described in terms of more demanding accounts of trust, i.e., accounts of trust that exceed rational expectations. However, both developers and the broader blockchain community are dedicated to the development of governance mechanisms that limit the capabilities of core-developers to assert power in an uncontrolled manner. Thereby, they reduce the vulnerability of users to developers. Thus, whether or not core developers are (or remain) as powerful as Walch portrays them continues to be a matter of debate.

Furthermore, the perspectives on users vary significantly in the literature on the role of trust in blockchain-based systems. While some inquiries only reflect on the information flow to and from the user and the user's set of possible actions, others also take into account how actual users are able to maneuver within systems. In a common practice in software development, "the 'trusted' label is given to systems that have been tested and proven to have met certain criteria" (Abdul-Rahman & Hailes 1998, p. 49). These criteria are usually technical. They pay less attention to the capabilities of actual users to maneuver within these systems. However, as Christopher (2016, p. 173) and Greenfield (2017) note, most users do not have the computer literacy necessary to understand and assess the code of their client applications, the blockchain protocol, or particular smart contracts they intend to use. Thus, even if users receive all relevant information necessary to verify certain actions by other users, such as the remittance of funds, they are most likely not able to assess them sufficiently. This phenomenon is not blockchain-specific. It rather emerges in the context of most of the use of modern-day technologies. Nickel (2013, p. 223) therefore points out that "[i]t is impossible for any one person […] to know enough about how technology works in these different areas to make a calculated choice about whether to rely on the vast majority of the technologies she/he in fact relies upon."

Therefore, most users must consult more computer literate human or institutional actors to assess the various technical subsystems on their behalf as, e.g., computer scientists who assess smart-contract code or cryptocurrency-wallet providers who support the administration of on-chain assets such as cryptocurrency tokens. These actors are not considered in the technical modeling, which largely assumes idealized users who can assess the information given to them. In order to leverage the trust-free or trust-minimizing features of the system that should allow users to not depend on trusting transactional counterparts and third-party intermediaries in a more demanding sense, most users cannot avoid making themselves vulnerable to new actors whom they need to trust in turn.

Thus, there necessarily are limits to the trust-minimizing features of blockchain-based systems. These features have an effect within the boundaries of the narrower technical system but do not fully extend it to actors at the fringes or outside of these boundaries or transactions that transcend them. However, the outlined manifestations of dependencies and vulnerabilities at the fringes of blockchain-based systems are contingent. For instance, the form and importance of these manifestations depend on the social permeation of blockchain-based systems. If cryptocurrency tokens are more widely accepted, the need to cross the boundaries of the system frequently, e.g., to exchange tokens back and forth into Fiat money, could vanish. Moreover, the relevance of cryptocurrency exchanges—one of the most significant factors of uncertainty—could accordingly be diminished substantially.

In this respect, both perspectives have a *raison d'être*. On the one hand, from an engineering perspective, it is reasonable to ignore these contingent factors at the fringes and consider only the technological features that can be impacted by employing means of the discipline. Here, Hawlitschek et al. (2018, p. 59) identify the trust-free properties that also Nakamoto and others describe as manifesting "as long as [the blockchain] operates as a closed ecosystem within its technical boundaries." On the other hand, it is crucial to recognize that—as Filippi (2018) points out illustratively in the title of one of her articles—"No Blockchain Is an Island." The salient vulnerabilities at the fringes of blockchain-based systems that challenge the depiction of them being trust-free or only based on trust in cryptographic algorithms are worth investigating, especially since they are a significant factor hindering widespread adoption of blockchain-based systems.

5 The Depictions of the Role of Trust in the Light of these Findings

As shown, the role of trust in blockchain-based systems can be characterized in various ways depending on which account of trust is applied and which components of the socio-technical system are considered essential. This section outlines the relevance of these decisions for the characterization of blockchain-based systems as being trust-free, being based on trust in technological components like cryptographic algorithms, or being based on distributed trust.

To highlight trust-minimizing features of the technology, blockchain advocates often apply a more demanding account of trust that sets higher prerequisites for a stance to be qualified as trust than just being based on rational expectations towards the trustees' behavior. Their positive notion of blockchain as a "trust-free" technology is only comprehensible if "trust-free" refers to the absence of the need to assess specific motivational factors of trustees as, e.g., the benevolence towards the trustor. Since the consensus mechanisms specified in blockchain protocols are based on an incentive system grounded in game theory, they need to be regarded as trust enhancing rather than trust-free if a less demanding account of trust based solely on rational expectation is applied. The notion of trust inherent in alleging that the blockchain technology is trust-free therefore must be read in terms of more demanding accounts of trust. The given incentive systems presuppose that actors are assessed regarding presumed self-interest only (Werbach 2018, p. 154). Goodwill, benevolence, or encapsulated interests are not taken into account, as required by more demanding accounts of trust. Within this framework, the innovation behind blockchain can be summarized as allowing users to move from having to *trust* institutional actors in a more demanding sense to only having to make *judgments of reliability* of actors based on game-theoretical assumption within a technologically predefined setting.

However, while the depiction of a trust-free technology appears to be generally tenable within this framework, it has been (over-)stretched by blockchain aficionados. "With a zeal bordering on the religious" they "trumpeted the trustlessness"[4] (Christopher 2016, p. 141) of the systems and declared it "one of the system's core virtues" (Christopher 2016, p. 172). By neglecting both limitations of the characterization, the application of a more demanding account of trust, as well as a focus limited to the actors within the narrow technical system, they hype an untenable image of a technological system that frees users from most

[4] Christopher (2016) uses the term "trustless" synonymously with how the term "trust-free" is used in this paper.

uncertainties and vulnerabilities without recognizing the emergence of new ones. Because some vulnerabilities of the users towards other actors are readily apparent, this narrative gets challenged. Critics here point to actors outside the boundaries of the narrower system. These include regulators (de Filippi and Wright 2018) and (core-)developers (Walch 2019) who can wield power individually, cryptocurrency exchanges which tend to fall victim to attacks regularly (Chohan 2018), and the users themselves who often lack the computer literacy necessary to navigate safely within the broader environment of blockchain-based systems (Christopher 2016; Greenfield 2017).

However, many of these critiques against the depiction blockchain-based systems as being trust-free apply an account of trust that is less demanding. They do not consider motivational factors like goodwill or encapsulated interest a prerequisite for genuine trust. Thus, ultimately, they do not distinguish between concepts like trust and mere judgments of reliability and adhere to an account that considers only positive predictive expectations. Sometimes, such a rational-expectation-based account of trust is even made explicit (cf. Botsman 2017). Because it sheds light on existing uncertainties and vulnerabilities of users in the broader environment of these systems, this critical counter-narrative against the exaggerated claims outlined in the previous paragraph is important and well-founded. Yet, it operates with divergent assumptions and is based on a terminology that differs from the ones used in *more reflective* depictions of blockchain as a technology that enables trust-free transactions.

The second depiction of the role of trust in blockchain-based systems suggests that users only have to place trust in algorithms, code, or math. The idea here is that trust is redirected from one trustee to another—from human and institutional actors to allegedly more trustworthy technological artifacts. However, trust in such entities is conceptually very different than interpersonal trust. Thus, suggesting that users are redirecting the same stance—trust—from human actors to technology (or concepts underlying these technologies) neglects that elements of interpersonal trust according to more demanding accounts cannot be modeled on these entities. They lack essential features such as the capacity for goodwill or encapsulated interest. Furthermore, reactive attitudes such as feeling betrayed in case of failed trust cannot be appropriately directed at technological artifacts. Therefore, speaking of trust in these entities is only meaningful if a less demanding account of trust is applied. However, such accounts of trust in technological artifacts or systems exist already (see Nickel 2013). To make the argument that "a new form of 'algorithmic trust' is created, one that significantly distinguishes itself from the more traditional typology of trust that was initially only between human agents" (Swan and Filippi 2017, p. 605), it is necessary to elaborate how this allegedly new form of trust differs from these more generalized notions of trust in technology. So far, this has not been addressed in the respective elaborations.

The third depiction suggests that the blockchain technology makes it possible to replace trust in individual actors with distributed trust. In this line of thinking, the technological components of the system do not constitute the trusted entity. Instead, they enable users to distribute trust over networks of actors without necessarily trusting any individual actor within the network based on given contextual features. Thus, this depiction considers both technical as well as human elements of the socio-technical system. By highlighting the distributed nature of the trustee, it allows distinguishing between the configuration of trust-based relationships in blockchain-based systems on the one hand and a centralized configuration of trust-relationships arranged around an intermediary or a central authority on the other hand.

Other setups containing distributed trust are already familiar from contexts such as accounts of trust in markets, accounts of trust in the wisdom of the crowd, or accounts of trust in reputation systems. These related concepts provide a basic understanding of how the term "distributed trust" can be understood as a meaningful concept, even though the feature of being distributed is incongruous with at least more demanding accounts of interpersonal trust. By pointing at the technological components of the socio-technical system that facilitate the distribution of trust, i.e., the cryptographic algorithms that enable the underlying consensus mechanism, this depiction also gives a clear picture of the innovation behind the blockchain technology. It suggests that the blockchain technology allows, on the one hand, distributing trust where it was hitherto not possible and, on the other hand, distributing trust by means other than those familiar from other contexts.

However, this depiction is only compatible with rather limited accounts of trust, too. The alleged distributed nature of trust does neither allow for the attribution of motivational factors as goodwill or encapsulated interest nor for the plausible application of reactive attitudes such as feeling betrayed in case of failure or breakdown of trust. Botsman (2017) and others make the adherence to a less demanding account of trust explicit by defining trust as, e.g., a "confident relationship with the unknown." Furthermore, the idea that this setup replaces trust in individual actors is also only tenable within the confines of technical modeling in which actors like (core-)developers are not considered. The role of these actors who have proven to currently have the capability to exert power individually remains unaccounted for in these considerations.

In addition, in their assessment of the role of trust in blockchain-based systems, some scholars (cf. Botsman 2017; Werbach 2018) consider not just one but multiple of the depictions introduced in this paper. The strength of these elaborations lies in that they give an overview of the newly established relationships and structural assurances as well as persisting and emerging vulnerabilities within the broader environment of blockchain-based systems. However, this openness often comes at the cost of conciseness. It subsumes very different conceptual stances under the umbrella term "trust." The stances that users have towards a distributed network of miners, towards contractual counterparties, and towards the underlying technical infrastructure vary greatly, even though dealing with uncertainty plays a role in all of them. By applying a very inclusive account of trust that solely focuses on dealing with uncertainty, they lose the conceptual framework to shed light on the differences between the various stances regarding, e.g., what characteristics of the respective trustees trustors are assessing and which moral dimensions some of the relationships might have.

6 Conclusion

Many scholars agree that one of the exceptional features of the blockchain technology lies in the unusual requirements it sets for users to place trust in other entities within the system when using it. There are undoubtedly various ideas on how the role of trust in blockchain-based systems differs from the role of trust in other setups. As shown, the ambiguity of the term "trust" plays a crucial role here. While most elaborations simply expect the underlying terminology to be self-explanatory, the implicitly underlying accounts vary greatly. Nevertheless, even though some depictions of the role of trust in blockchain-based systems appear to be incongruous with others, most of them cannot be rejected offhand. In a benevolent interpretation, there are accounts of trust that allow

all the entities suggested as trustees at the core of the setup to be covered meaningfully. However, some of the depictions contribute more to the understanding of how blockchain-based systems work and what the critical issues are. Especially the depiction of blockchain as a technology that enables technologically facilitated trust in a distributed network of actors that are not trusted individually can be positively highlighted here. Contrary to other depictions, it allows considering both technical components as well as key actors simultaneously.

The heterogeneous terminologies in the respective lines of argumentation make it increasingly challenging for scholars referencing one another in the academic discourse. Propositions that are made—at least implicitly—based on one account of trust are often attacked from a perspective based on a different conceptual and terminological basis. The most striking example here is the debate on whether or not the blockchain technology enables trust-free transactions. As shown, a reasonable argument can be made for this within the limited purview of technical modeling and based on a more demanding account of trust. However, critics of this notion as well as blockchain aficionados often ignore the limitations of this argument and treat the idea of trust-free transactions as an alleged system feature that without more ado can be leveraged in the everyday usage of implemented systems. Here, the lack of a shared terminology contributes to both an unwarranted critique of the initial argument and an unwarranted hype surrounding alleged features the blockchain technology.

In spite of this, the various affected scholarly disciplines and traditions of thought do not allow a uniform underlying account of trust to be defined. To tackle these issues nevertheless, Hardin's (2002, p. 87) suggestion that discourse participants need to specify their terminology regarding trust should be taken more seriously. Thus, since a shared terminology across the multitude of involved disciplines does not appear to be an attainable goal, it is paramount that scholars reflect on divergent views and make underlying assumptions and concepts explicit. In the scholarly context, this should be considered both in the formulation and presentation of arguments as well as in academic evaluation processes.

For this purpose, scholars can fall back on comprehensive works that provide overviews over various accounts of trust in a more general sense (Hardin 2002; McLeod 2006; Simon 2013) and on accounts of trust in specific entities and domains. In the context of the discourse on blockchain, e.g., interpersonal trust (Baier 1986; Hardin 2002), trust in groups and organizations (Hawley 2017), trust in technological systems (Nickel 2013), and trust in game-theoretical settings (Gambetta 1988; Voss and Tutic 2020) are particularly noteworthy. Here, future research could build on the findings of this paper by providing scholars with a taxonomy of the accounts of trust relevant in blockchain research.

If these overviews do not provide accounts that suit a specific proposition or argument, e.g., in cases where scholars allege that blockchain establishes a *new* form of trust (cf. Swan and Filippi 2017; Werbach 2018), it is nevertheless necessary to introduce an explicit definition of trust. Here, the fact that trust is a reductive term can be utilized as a starting point for the explication of hitherto implicit assumptions on the nature of trust. Because trust "is reducible to other things that go into determining trust" (Hardin 2002, p. 57), these "other things" or components can be pointed out individually. For instance, in the case of the depiction of blockchain as a technology that allows for trust-free transactions, this approach requires clarifying that it is the consideration of specific motivations—which allegedly are prerequisites of trust—that become negligible within the boundaries of the narrower technical system.

Depending on the degree of divergence from existing accounts of trust, these new conceptions are also a worthy subject for philosophical investigations and further research. Particularly the notion of distributed trust in the relationship of users and miners enabled through specific technical features which establish an economic incentive system appears to be noteworthy in this regard. Based on the works of scholars such as Antonopoulos (2014 2017), Hawlitschek et al. (2018), and Werbach (2018), who describe the relationship of trust and technical features in more detail, comparisons to adjacent phenomena such as trust in markets can be drawn to develop an analogous concept. These novel accounts would complement the aforementioned taxonomies.

Acknowledgments I acknowledge financial support from Hamburg Ministry of Science, Research and Equality in the project Information Governance Technologies under the reference LFF-FV 34.

Funding Information Open access funding provided by Projekt DEAL.

References

Abdul-Rahman, A., & Hailes, S. (1998). A distributed trust model. In *Proceedings of the* 1997 Workshop on new security paradigms. Symposium conducted at the meeting of ACM.

Antonopoulos, A. (2014). Bitcoin security model: Trust by computation. *Forbes. com, February*, *20*. Retrieved from http://radar.oreilly.com/2014/02/bitcoin-security-model-trust-by-computation.html

Antonopoulos, A. (2017). *Mastering Bitcoin: Programming the open Blockchain* (Second ed.). Sebastopol, CA: O'Reilly Media.

Atzori, M. (2015). Blockchain technology and decentralized governance: Is the state still necessary? *SSRN Electronic Journal. Advance online publication.* https://doi.org/10.2139/ssrn.2709713.

Baier, A. (1986). Trust and antitrust. *Ethics, 96*(2), 231–260. https://doi.org/10.1086/292745.

Beck, R., Czepluch, J. S., Lollike, N., & Malone, S. (2016). Blockchain-the gateway to trust-free cryptographic transactions. In *ECIS*.

Botsman, R. (2017). *Who can you trust? How technology brought us together and why it might drive us apart (first edition (eBook))*. New York: PublicAffairs.

Buterin, V. (2014). DAOs, DACs, DAs and More: An Incomplete Terminology Guide. Retrieved from https://blog.ethereum.org/2014/05/06/daos-dacs-das-and-more-an-incomplete-terminology-guide/

Chohan, U. (2018). The problems of cryptocurrency thefts and exchange shutdowns. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.3131702.

Christopher, C. M. (2016). The bridging model: Exploring the roles of trust and enforcement in banking, Bitcoin, and the blockchain. *Nevada Law Journal, 17*, 139.

DuPont, Q., & Maurer, B. (2015). Ledgers and law in the Blockchain. *Kings Review (23 June 2015)* http://kingsreview.co.uk/magazine/blog/2015/06/23/ledgers-and-law-in-the-blockchain.

Filippi, P. de. (2017). In Blockchain we Trust: Vertrauenslose Technologie für eine vertrauenslose Gesellschaft. In Rudolf-Augstein-Stiftung (Ed.), edition suhrkamp: Vol. 2714. Reclaim Autonomy: Selbstermächtigung in der digitalen Weltordnung (pp. 53–81). Berlin: Suhrkamp.

Filippi, P. de (2018). No Blockchain Is an Island. Retrieved from https://www.coindesk.com/no-blockchain-island/

de Filippi, P., & Wright, A. (2018). *Blockchain and the law: The rule of code*. Cambridge: Harvard University Press.

Freeman, S., Beveridge, I., & Angelis, J. (2020). Drivers of digital trust in the crypto industry. In M. Ragnedda & G. Destefanis (Eds.), *Routledge studies in science, technology and society. Blockchain and Web 3.0: Social, economic, and technological challenges* (pp. 62–77). London: Routledge.

Gambetta, D. (1988). Can we trust trust? In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 213–237). Oxford: Blackwell.

Greenfield, A. (2017). Radical technologies: The design of everyday life. London, New York: Verso.

Hardin, R. (2002). Trust and trustworthiness. The Russell Sage Foundation series on trust: Volume 4. New York: Russell Sage Foundation. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=1069635

Hawley, K. (2017). Trustworthy groups and organizations. In P. Faulkner & T. Simpson (Eds.), *The Philosophy of Trust (pp. 230–250)*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198732549.003.0014.

Hawlitschek, F., Notheisen, B., & Teubner, T. (2018). The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy. *Electronic Commerce Research and Applications, 29*, 50–63. https://doi.org/10.1016/j.elerap.2018.03.005.

Hoffman, R. (2014). The Future of the Bitcoin Ecosystem and "Trustless Trust": Why I Invested in Blockstream. Retrieved from https://www.linkedin.com/pulse/20141117154558-1213-the-future-of-the-bitcoin-ecosystem-and-trustless-trust-why-i-invested-in-blockstream

Holton, R. (1994). Deciding to trust, coming to believe. *Australasian Journal of Philosophy, 72*(1), 63–76. https://doi.org/10.1080/00048409412345881.

Jones, K. (1996). Trust as an affective attitude. *Ethics, 107*(1), 4–25. https://doi.org/10.1086/233694.

Krishna, A. (2015). Blockchain: It Really is a Big Deal. Retrieved from https://www.ibm.com/blogs/think/2015/09/blockchain-really-big-deal/

Mallard, A., Méadel, C., & Musiani, F. (2014). The paradoxes of distributed trust: Peer-to-peer architecture and user confidence in Bitcoin. *Journal of Peer Production*. (4), 1–10.

Maurer, B., Nelms, T. C., & Swartz, L. (2013). "When perhaps the real problem is money itself!": The practical materiality of Bitcoin. *Social Semiotics, 23*(2), 261–277. https://doi.org/10.1080/10350330.2013.777594.

McLeod, C. (2006). Trust. Retrieved from https://plato.stanford.edu/entries/trust

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. Retrieved from https://bitcoin.org/bitcoin.pdf

Nickel, P. J. (2013). Trust in Technological Systems. In M. J. Vries, S. O. Hansson, & A. W. M. Meijers (Eds.), *Philosophy of engineering and technology*, *Norms in technology* (Vol. 9, pp. 223–237). Dordrecht: Springer.

Simon, J. (2013). Trust. In D. Pritchard (Ed.), *Oxford bibliographies*. New York: Oxford University Press. https://doi.org/10.1093/obo/9780195396577-0157.

Swan, M. (2015). *Blockchain: Blueprint for a new economy*. Beijing: O'Reilly.

Swan, M., & de Filippi, P. (2017). Toward a philosophy of Blockchain: A symposium: Introduction. *Metaphilosophy, 48*(5), 603–619. https://doi.org/10.1111/meta.12270.

The Economist (2015). The trust machine. Retrieved from https://www.economist.com/leaders/2015/10/31/the-trust-machine

Underwood, S. (2016). Blockchain beyond bitcoin. *Communications of the ACM, 59*(11), 15–17. https://doi.org/10.1145/2994581.

Velasco, P. R. (2017). Computing ledgers and the political ontology of the Blockchain. *Metaphilosophy, 48*(5), 712–726. https://doi.org/10.1111/meta.12274.

Voss, T., & Tutic, A. (2020). Trust and game theory. In J. Simon (Ed.), *The Routledge handbook of trust an philosophy*. Routledge.

Walch, A. (2019). In code(rs) we trust: Software developers as fiduciaries in public Blockchains. In I. Lianos, P. Hacker, S. Eich, & G. Dimitropoulos (Eds.), *Regulating Blockchain: Techno-Social and Legal Challanges* (pp. 58–81). Oxford University Press.

Werbach, K. (2017). Trust, but Verify: Why the Blockchain needs the law. *Berkeley Technology Law Journal.*

Werbach, K. (2018). *The blockchain and the new architecture of trust. Information policy series*. Cambridge, Massachusetts, London, England: The MIT Press.