Strategies to Tackle Disinformation: Operationalizing Zero Trust

Allison Wylde
Data Science for Common Good Research Group, Glasgow Caledonian University, London, UK

allison.wylde@gcu.ac.uk

Abstract: Disinformation is now acknowledged as one of the leading threats to global security. Although trust is a central foundation in the take-up of disinformation and in its resulting loss of trust, little is known about the mechanisms of trust. In response, this conceptual paper first reviews a specific strand of trust and distrust literature from management, organization and conflict management studies models to attempt to disentangle the trust issue in disinformation. The method employed was based on a purposive literature review. This approach allowed generating a deep understanding of the foundational literature, in the context of understanding trust in disinformation and a transformative approach from cybersecurity zero trust as a potential solution to operationalize the aims of this research. Drawing from the emerging findings from the review, the paper then proposed leveraging zero trust as a tactic to counter disinformation. Although the limitations of a purposive literature review approach are acknowledged, calls for further research and action are presented thereby helping bridge potential methodological issues. The contribution of this paper presents an early-stage framework setting out the key tactics involved in operationalizing and achieving a zero trust mindset to safeguard against disinformation. Key implications for government, defense practitioners, academics and stakeholder communities are discussed.

Keywords: Disinformation, Trust, Zero trust, ZT, Presumptive trust, Non-presumptive trust

1. Introduction

One unanswered conundrum in disinformation concerns trust; disinformation spreads because it contains a grain of truth, prompting individuals to trust the content (DeFranco et al., 2021). Yet, disinformation erodes trust (Fallis, 2015); examples include reduced public trust in governments during the Covid-19 pandemic and in elections (DeFranco et al., 2021). Yet, little is understood about the processes involved in countering trust acceptance and the erosion of trust in the context of disinformation.

One promising approach is zero trust (ZT); this practice has emerged as a transformative approach mandated for cybersecurity (The White House, 2021: DoD, 2022). Although ZT offers clear benefits as a mechanism for cybersecurity, to date to the best of the author's knowledge, studies of trust and ZT in disinformation have received limited attention.

This paper teases out distinctive features of the processes involved in assessing trust, distrust and, finally, ZT. At this early stage of research, a context based method (Denzin and Lincoln, 2011), limited in scope to a review of specific literature was undertaken, including; trust and distrust and the universal sequence (USQ) model of building trust and distrust (Dietz, 2011: Six and Latusek, 2023). The aim was to lay a theoretical foundation to address both gaps in the literature and the lack of consistency in studies to date.

After the introduction, section 2 explores trust and the trust assessment, accounting for who or what is being trusted to do what, in the context of disinformation. Section 3 provides an argument that the USQ framework can be drawn on to understand trust, the trust assessment and ZT in disinformation. Section 4, sets out the new framework and tactics. The final section, 5, presents the conclusion, implications, contribution, limitations and promising directions.

2. Trust, ZT and Disinformation

Although trust is in daily use, there is a lack of a common understanding of trust. This conceptual paper leverages well-developed theory, the universal sequence for building trust (Dietz, 2011) and building distrust (Six and Latusek, 2023) applied to ZT. What follows is limited to disentangling the important processes and concepts essential in addressing gaps in our understanding of trust and ZT in the context of disinformation; discussed next.

2.1 Trust

The aim of this paper centres on understanding trust as a subjective and relational construct (Marsh, 1994, 33) that exists between agents (human, non-human, organizational and institutional) and is achieved through decision-making under uncertainty (Mayer et al., 1995: Rousseau et al., 1998). Thus, views of trust scholars from management, conflict management and organizational studies provide a focused and well-established base for research.

Trust is founded on positive expectations and a willingness to accept vulnerability on the part of trustors (Mayer et al., 1995) under uncertainty (Rousseau et al., 1998), moderated by prior experiences of trust (Deutsch, 1958). In addition, trust encompasses several levels of relations, interpersonal trust, organization-level trust and institutional trust and at levels of low to high trust (Fulmer and Gelfand, 2012).

Trust has been studied in a range of contexts, including trust in technology or trust in governments, with some policy researchers seeing trust as bound with power. Therefore, in seeking to address the context here, the definition that serves for this paper draws together theory from the integrative trust model (ITM) (Mayer et al., 1995), including prior experiences of trust (from the conflict management literature, notably, Deutsch, 1958) while accounting for the level of trust in different entities or agents (Fulmer and Gelfand, 2012).

Trust is thus defined as the willingness of a trustor to be vulnerable to the actions of a trustee (Mayer et al., 1995) based on uncertainty (Rousseau, et al., 1998) alongside confident positive expectations of that trustee (Mayer et al., 1995), together with a trustor's experience of and ability to trust (Deutsch, 1958). The central idea is that the trust follows a universal sequence, comprising a set of inputs together with beliefs, decisions and actions involved in building trust (Dietz, 2011) and distrust (Six and Latusek, 2023). ZT processes are reviewed next.

2.2 Zero Trust (ZT)

To the best of the author's knowledge, doctoral research published in 1994 was the first to use the term ZT (Marsh, 1994). The idea was that ZT was a formalized decision not to trust based on a rationale that there was no knowledge of, or there was indifference to, an entity or there was a prior negative experience (Marsh, 1994, 75). ZT then became popularized in cybersecurity, due to issues including a blurring of organizational boundaries where third parties were found to be inside organizational supply chains, giving rise to views that networks should be considered as breached, and that trust is a vulnerability (Kindervag, 2010).

Next, standards for ZT emerged. NIST views ZT as a designed to minimize uncertainty in enforcing least privilege per-request decision (Rose et al., 2020). The UK's NCSC's guidance called for continuous authentication, verification and authorization to counter the prevalence of trust inside networks, which allows malevolent actors to move laterally (NCSC, 2021). The standards for ZT focus on the verifications of identity, for individuals (users), or devices, or software (NCSC, 2021). NIST proposed a risk-based approach involving minimizing uncertainty and enforcing decisions based on continuous authentication and authorization (Rose et al., 2020). As discussed in Section 1, ZT was then enforced as a government requirement (The White House, 2021) and for defense (2022a). Recent concerns regarding the slow progress in ZT implementation are examined in sections 3 and 4.

The question of presumptive trust is at the heart of this paper. In a similar though different context, as you start your car or open your fridge, do you trust that the car will start, and that the fridge will have kept food cold? What happens under ZT? For parsimony, this paper is framed as questions of; how disinformation may be assessed under ZT? (Wylde 2021; 2022).

2.3 Disinformation

Key thinking on the definitions in disinformation are explored next. Interestingly, while some of the central terms are agreed, some key differences remain in definitions across jurisdictions.

The US Cybersecurity and Infrastructure Agency (CISA, nd.) defines disinformation as "deliberately created to mislead, harm, or manipulate a person, social group or country" (CISA, nd.). In the EU, disinformation refers to "verifiably false or misleading information that is created, presented and disseminated for economic gain or intentionally to deceive the public and may cause public harm" (European Commission, 2018). These definitions possess shared aims of creating harm through the practices of deception and manipulation. Yet the scale varies, with the EU focused on the "public" compared with a broader scope from the US, ranging from the "person to the social group and country" (CISA, nd.). As the EC includes the notion of a beneficiary through economic gain, in one sense, the scale of a "country" could be viewed as the scope of the realm of "country" economics (European Commission, 2018).

Following the definitions, researchers agree that disinformation involves processes that are "deliberate and often orchestrated attempts to confuse or manipulate people" (Ireton, 2018). Further clarity is provided by ideas that disinformation can be defined as based on the presence (simultaneously) of three important criteria, the presence of malign actors whose identity is masked, the release of information that has intentionally harmful or destructive content, and, a predetermined political, military, economic or social objective (Murphy, 2023).

Strategies employed by malevolent actors aim to increase the impact (potency) of disinformation through including some sources of information that may accurate (or fake) together with information that may be false or fake (Arce, 2024). Studies have found that false or fake information is less likely to be challenged, when compared with false sources, largely moderated by a target's own willingness to believe alignment with their own world view.

Although trust has been referred to in disinformation research, for example, through disinformation involving the active engagement of someone to mislead, with the resulting harm indirectly "eroding trust" (Fallis, 2015). Indeed, although trust is variously considered as a factor in disinformation, the role of trust is underexplored. One strand of thinking of trust in disinformation says as a society we need to trust the internet less (De Franco et al., 2021). For this paper the focus is on the aspect of trust in disinformation.

Current recommendations for interventions include: improving education for users to help them prioritise quality over clicks; AI and ML to validate and rate content; facilitating individuals' preferences for accuracy; building trust; and, promoting interoperability (through blockchain) allowing users moves between platforms (De Franco et al., 2021).

Summing up, this section has aimed to tease out key thinking on trust, trust assessment, ZT and disinformation. Although trust, trust assessment, ZT are well specified, the situation is less clear for disinformation. Current differences in definitions and scale of impact for disinformation have arguably resulted in barriers to the take-up of common strategies. Understanding the application of ZT as a disinformation intervention is the central aim of this paper, and as a starting point the processes involved in interpersonal (inter-agent) trust are re-examined, to create a robust framework as the basis for a decision to trust/ or not to trust. The rationale is based on viewing a decision to trust made at the level of an individual trustor. As such, it is a trustor who decides whether to trust or not, or to distrust an individual or an entity, or indeed to apply ZT. Trust is now discussed, drawing on prominent theories of trust.

2.4 The Universal Sequence (USQ): Trust and/or distrust (T-DT)

First conceptualized by Dietz (2011), drawing from the ITM, the USQ involved in trust and distrust is viewed as founded on a five-stage process comprising: inputs, beliefs, decision, actions, and feedback (Dietz, 2011: Six and Latusek). Each trust (distrust) encounter is viewed as a separate process, so for trustors each experience occurs as single experience (Dietz, 2011). The USQ process incorporating distrust is discussed next (Six and Latusek, 2023).

- Stage 1 of the USQ, Inputs, five separate types of input are involved (IN1-IN5); in the condition of trust, all types of inputs (T-IN1-5) are trusted, whereas under conditions of distrust (DT-IN1-5), all the inputs are distrusted.
- Stage 2, Beliefs (T-B1); in trust, as with stage 1, beliefs are assumed as trustworthy, and belief (T-B2); is based on confident positive expectations. The reverse holds true for DT, with (DT-B1); being untrustworthiness and (DT-B2); confident negative expectations.
- Stage 3, Decisions, if trust is held (T-D1); then the decision is a willingness to render oneself vulnerable, with the opposite for DT, (DT-D1); intention not to be vulnerable.
- Stage 4, Actions are considered, in trust (T-A1); the action involves risk-taking behaviour, and the reverse for DT (DT-A1); with based on no action.
- Stage 5, the Feedback loop for trust (T-FBL); feedback is positive and based on experiences of 'trusting' with the reverse for DT (DT-FBL1); feedback is protective and based on experiences of 'distrusting', hence there is little or no interaction.

2.5 The Universal Sequence (USQ): ZT (ZT)

Turning now to ZT, for simplicity the discussion presents a view from the perspective of a decision to adopt a ZT stance in an individual trustor- relationship, for example, as ZT applied to an individual unit of information. Although the processes involved in ZT act in parallel (and do not follow a sequential approach), for simplicity the separate elements are discussed separately, by adapting the sequences of the USQ model. The separate ZT stages 1-5 are considered in turn, starting with Stage 1, and the five Types of Inputs.

- Stage 1, Input Type 1, input based on the trustors' propensity (ZT-IN1); for ZT (at ZT).

Stage 1, Input Type 2, input is the assessment of a unit of information (ZT-IN2); under ZT input networks are assumed to be breached. In addition, in ZT a unit of information may be judged as possessing motives that are

malevolent and may result in bad behaviour. For the ABI part of assessment, a unit of information may be viewed as possessing either a good or a bad ability. As an example, a good ability may be necessary to breach the network, with a bad purpose. This good ability could also encompass a negative (bad) benevolence and/ or a bad integrity. In a reverse state, if the unit of information is judged as possessing a bad ability, then in this scenario an indication of a bad ability could result in the trustor detecting a breach (the presence of disinformation). In ZT, the assessment output may or not inform the trustor's beliefs (ZT-BI) in stage 2 of the USQ.

Stage 1, Input Type 3, input nature of relationships (ZT-IN3); relationships viewed as ZT.

Stage 1, Input Type 4, the input focuses on domain-specific concerns (ZT-IN4); for ZT, the domain-specific concerns (ZT-IN4); involve active measures for cybersecurity, compared with in T&DT-IN4, where the domain-specific concerns are not specified.

Stage 1, Input Type 5, the inputs focus on the institutional context (ZT-IN5); in ZT includes active measures for cybersecurity in high-risk environments and/ or organizations.

- Stage 2, Beliefs (ZT-B1); in ZT beliefs are the same as (DT-B1), based on untrustworthiness (incompetence, malevolence and deceit and negative ABI) and for (ZT-B2), again, the same as (DT-B2), confident negative expectations. Under conditions of ZT, a third belief (ZT-B3), breached networks (disinformation), hence continuous monitoring.
- Stage 3, Decisions (ZT-D1); for ZT decisions consistent with (DT-D1), no vulnerability.
- Stage 4, Actions (ZT-A1); for ZT action is do not take risk, in (ZT-A2); action is continuous assessment and monitoring, in ZT an additional action, (ZT-A3); in ZT, a third action involves granting trust or not granting trust.
- Finally, in Stage 5, The Feedback Loop,(ZT-FB1); under ZT feedback assumes that networks are breached, for ZT an additional feedback loop is included (ZT-FB2); continuous feedback based on continuous assessment and monitoring.

In ZT the stages in trust/ distrust (as above) are viewed as separate and not interconnected, thus the dimensions have no influence on any other dimension, indeed, they are viewed as acting in parallel. The assessment and decision-making of ZT following the USQ approach are now considered in further detail under conditions of ZT. The goal of ZT is twofold, firstly, to undertake the correct assessments and decisions and secondly to prevent any mistakes, for example, during both the assessment and decision processes involved in granting/ or not-granting trust. Granting trust is conditional and relies on the identity of a unit of information being correctly verified. Any bias may result in including a trustee's character or motives and evaluating the nature of the relationship (dimension 1 type 2 and 3, ZT-IN2 and ZT-IN3).

There are several points of overlap and simultaneous assessment in the USQ. This means that during several stages of assessment and decision-making, trustors must be aware of, and supersede, their personal disposition to trust (ZT-IN1), and at the same time, recognize and overcome their personal beliefs of trustworthiness (ZT-IN2 and ZT-B1). This self-awareness is essential to correctly understand and thus evaluate inputs and make decisions. Given that simultaneous and parallel tasks require the involvement of personal beliefs, trustors may become overwhelmed. Also, under conditions of ZT, an additional belief is present, that networks are breached, which may be at odds with trustors personally held beliefs.

By further specifying the various conceptual categories and dimensions in the USQ, the processes, from inputs to outputs under conditions of ZT, have been examined in granular detail. In addition, from the perspective of examining an individual ZT entity, the USQ has enabled a selective interpretation, and then justification as to how to assess a particular unit of information (as an identity) to determine if trust can be granted. The processes elaborated here may aid detection and help flag the presence of a malevolent actor and/ or a breached network.

As a final note to this section, under conditions of ZT the importance of considering the role of the domain-specific concerns and the institutional context are highlighted. For ZT, the primary concern and context are cybersecurity, compliance, and high-risk settings. Indeed, for the successful operation of ZT, the central context is the belief of an active network breach; ZT thinking could also be drawn on to help the processes of managing incidents and/or crises.

3. Discussion: Towards a Shift, Trust and ZT

Although the goals of implementing ZT are grounded in ideas that ZT is founded on non-presumptive trust and in a risk-based approach to trust (Rose et al., 2020), to date the literature has been slow to characterize the validation processes in ZT (Wylde, 2021; 2022). This paper now re-examines ZT validation as a preface to operationalizing ZT.

As a starting point, recent examples of ZT in policy are re-examined to determine the applicability of the USQ model in helping operationalize ZT. As set out in Section 3, the USQ models specify the processes for assessing trust, distrust and ZT based on separate stages, each viewed as a discrete and on a linear pathway (Wylde, 2021). However, in practice, trust relations are more complex and dynamic: How far do I trust? What if I trust one and not the other at the same time? Indeed, it also appears, in the context of disinformation, that a target's willingness to believe may be related to their experiences of trusting (Mayer et al., 1995).

As discussed, in practice the adoption of ZT has been limited by issues such as the presence of legacy systems, transparency in cloud providers and a workforce that may be reluctant to use new processes (Golden et al., 2021). Barriers such as a lack of consistency across ZT management practices and systems may also impact operational efficiency (Yeo et al., 2023: Wylde, 2021). In warfare, ZT appears promising for the protection of high-risk settings.

Returning to ZT directives, the DoD ZT capabilities model sets out 152 activities to be achieved for ZT, with enablers ranging from; cross-cutting and non-technical capabilities, to those that address culture, governance and the execution enablers, ranging from doctrine organizational through to people, facilities and policy elements (DoD, 2022). It is likely that for each of the separate 152 activities and for their relations with one another, trust assessments and decision-making will be involved. Although AI offers automation capabilities that can deal with the scale of such an undertaking, the fine-tuning and calibration of trust assessments is critical. It is suggested that a framework such as the USQ could be further developed to help address issues in trust and ZT implementation. Modules such as the USQ could be automated through AI learning (machine learning) and knowledge gaps to provide processes that continuously monitor and evaluate trust and ZT - supporting ZT policy enforcement processes and policy engines (Rose, 2020). Finally, leveraging a model such as the USQ that recognizes the important role of the trustor's personal beliefs and predisposition could be an essential consideration in helping to build a workforce cybersecurity mindset (DoD, 2022b, 11). In sum up, as ZT itself is founded on continuous decision-making and as this paper has argued, ZT could be actioned as an intervention to disinformation.

4. Conclusion

This paper argues that a positive intervention in countering disinformation can be achieved through leveraging ZT, the USQ and trust, teased out in a disinformation context.

Several important differences in trust processes identified allowed for the separate and parallel activities involved in ZT to be better understood. Promising avenues for future research include work to further develop the USQ model, for example, during the input stage (USQ), trustors' ZT decision-making could be evaluated to determine the levels of presumptive and non-presumptive trust to determine the likelihood of impacts on their ZT assessments in subsequent decision-making in the face of potential disinformation. Further research could examine the implementation of ZT policy and subsequent assessments of trust in data-driven models using AI and simulation technology to model, test and validate operations. Future studies could consider how the USQ factors could be expanded to allow for further evaluation, validation and rating through ML or AI models (De Franco et al., 2021). AI models could be designed to integrate ZT security training alongside policies to help build a disinformation prevention culture. Finally, as the current practices for designing AI models are at risk of subversion, the role of human factors and the processes in trust and ZT could be considered during the development of AI and ML models for countering disinformation – for trust.

In all work, limitations exist. The USQ is acknowledged as based on separate, mechanistic and linear elements. As such, the USQ is fails to account for the non-linear dynamics of ZT. Building a shared understanding is relevant given the current lack of interoperability among governance, standards and vendors' systems prompting calls for further research.

Acknowledgements

Thank you to colleagues across the trust, cyber security and disinformation community for helpful discussions. An earlier version of this paper was submitted to CyCon24, thank you reviewers and organizers for your insightful feedback and suggestions.

References

Arce, D. 2024. Disinformation Strategies, *Defence and Peace Economics*, 35, no. 6, 659-672, viewed, 01 Feb. 2024, <https://doi.org/10.1080/10242694.2024.2302236>.

CISA. Nd. *Foreign Influence Operations and Disinformation*. viewed, 01 Feb. 2024, <https://www.cisa.gov/topics/election-security/foreign-influence-operations-and-disinformation>.

DeFranco, J. F., Kshetri, N. Sharma, R. & Rojas-Torres, D. 2021. Mitigating disinformation and building trust in social media. *IT Professional* 23, no. 6: 62-66.

Department of Defense. 2022. *DoD Zero Trust Strategy*. October 21, 2022, https://dodcio.defense.gov/Portals/0/Documents/Library/DoD-ZTStrategy.pdf>.

Denzin, N.K. and Lincoln, Y.S. eds., 2011. *The Sage handbook of qualitative research*. Beverley Hills, CA: Sage.

Deutsch, M. 1958. Trust and suspicion. *The Journal of Conflict Resolution* 2, no 4: 265-79.

Dietz, G. 2011. Going back to the source: Why do people trust each other? *Journal of Trust Research* 1, no. 2: 215-222.

European Commission. 2018. *Action plan against Disinformation*. Brussels, 5.12.2018, JOIN(2018) 36 final, viewed, 01 Feb. 2024, <https://op.europa.eu/publication-detail/-/publication/e18263d2-f962-11e8-8885-01aa75ed71a1>.

Fallis, D. 2015. What is disinformation? *Library trends* 63, no, 3: 401-426.

Fulmer, A. C. & Gelfand, M. J. 2012. At what level (and in whom) we trust: Trust across multiple organizational levels. *Journal of Management* 38, no. 4: 1167-1230.

Ireton, C. 2018. *Journalism, 'Fake News'& Disinformation: Handbook for Journalism Education and Training*. UNESCO Publishing.

Kindervag, J. 2010. No more chewy centers: *Introducing the zero trust model of information security*. Forrester Research, September 14, 2010, updated September 17, 2010, viewed, 01 Feb. 2024, <https://media.paloaltonetworks.com/documents/Forrester-No-More-Chewy-Centers.pdf>.

Lewicki, R. J., McAllister, D. J. & Bies. R.J. 1998. Trust and distrust: New relationships and realities. *Academy of Management Review* 23, no. 3: 438-458.

Marsh, S. P. 1994. *Formalising trust as a computational concept*. PhD. Diss., Stirling Uni.

Mayer, R., Davis, J.H. & Schoorman, D.F. 1995. An integrative model of organizational trust. *Academy of Management Review* 20, no. 3: 709-734.

Murphy, B. 2023. Evaluating the Ambiguous Cognitive Terrain: A Framework to Clarify Disinformation. *Journal of Information Warfare*, 22 no.3: 9-27.

NCSC. 2021. *Zero Trust Architecture*. July 23, 2021. viewed, 01 Feb. 2024, <https://www.ncsc.gov.uk/collection/zero-trust-architecture.

Rose, S., Borchert, O., Mitchell, S. & Connelly, S. 2020. *Zero trust architecture. NIST special publication*, August 2020, viewed, 01 Feb. 2024, <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf>.

Rousseau, D. M., Sim B., Sitkin, B. & Camerer. C.1998. Not so different after all: A cross-discipline view of trust. *Academy of Management Review* 23, no. 3: 393-404.

Six, F. E. & Latusek, D. 2023. Distrust: A critical review exploring a universal distrust sequence. *Journal of Trust Research* 13, no. 1:1-23.

The White House. 2021. *Executive order on improving the nations cybersecurity*. May 12, 2021, viewed, 01 Feb. 2024, <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/>.

Wylde, A. 2021. *Zero trust: Never trust, always verify*. International conference on cyber situational awareness, data analytics and assessment (CYBERSA), Jun 14, 2021: 1-4). IEEE.

Wylde, A. 2022. *Questions of trust in norms of zero trust*. In: Arai, K. (eds) Intelligent Computing. SAI 2022. Lecture Notes in Networks and Systems 508. Springer, Cham, viewed, 01 Feb. 2024, <https://doi.org/10.1007/978-3-031-10467-1_51.