Calibrated Trust as a Means to Build Societal Resilience Against Cognitive Warfare

Esther Kox, Neill Bo Finlayson
Julia Broderick-Hale, José Kerstholt
THE NETHERLANDS

esther.kox@tno.nl

neill_bo.finlayson@tno.nl

julia.broderick@tno.nl

jose.kerstholt@tno.nl

ABSTRACT

Intervening to bolster trust is critical to strengthening societal resilience and to reducing vulnerability to cognitive warfare threats. Trust, however, is a broad, multi-layered, dynamic construct that is difficult to grasp and to intervene upon. The suitability and potential effectiveness of an intervention may vary depending on the specific trust dynamic. Therefore, to select the optimal trust intervention for a given situation, one must understand the dynamics of trust. We describe trust as the ever-fluctuating product of ongoing trust calibration. The objective of this paper is to conceptualise the trust calibration cycle to help identify the status of a trust dynamic, provide foresight into future developments and to support selection of appropriate interventions. To promote understanding of what trust is and how it works, we first clarify the conceptual boundaries of trust, including mistrust and distrust, types of trust and the elusive ideal: calibrated trust. Our conceptualisation of the trust calibration cycle entails three phases in which a trust dynamic can be characterised: 'building', 'managing', and 'repairing'. The context and nature of these phases and exemplar interventions are discussed (using social science literature), in an effort to bring together theory and practice. Finally, implications, limitations and key takeaways are discussed. By reimagining the trust calibration cycle, we hope to support the ability of policymakers to understand, identify and intervene upon trust in society in order to bolster societal resilience as a means of mitigating the effects of cognitive warfare.

1.0 INTRODUCTION

The proliferation of digital and social media has manifested a new domain of conflict: cognitive warfare. The aim of cognitive warfare – a form of hybrid warfare – is to gain advantages over an adversary by reshaping and manipulating their understanding of military, political or social environments in favour of one's own strategic or tactical objectives [12]. In essence, the human mind becomes the battlefield and information is the weapon, with the aim being to change how a target population thinks and, in turn, acts [4], [10] It is a unique type of non-kinetic warfare that goes beyond just controlling flows of information and is more concerned with the ability to control how people react to information [6].

The long-term objective of cognitive warfare is often the disruption and destabilisation of enemy states, populations or societies. The idea is to sow seeds of distrust and division regarding governance, the rule of law or democratic processes, with the aim of stoking societal cleavages and polarisation so that "enemies destroy themselves from the inside out" [6]. In its most potent form, cognitive warfare "has the potential to fracture and fragment an entire society" [10]. This is because it aims at disrupting interpersonal relationships

by targeting human vulnerabilities, such as trust, on both the societal and the individual level [29]. Trust is a foundational element of a functional society and has been described as "the glue that makes dependencies and connections strong and healthy in democracies as well as supports the foundations of democratic system" [22] (p.32). It is precisely this dependency on trust that means it is also a vulnerability and one which malign actors seek to exploit.

Russia, for example, is a foremost proponent of cognitive warfare. The Kremlin's attempts to interfere in foreign elections are well publicised [34], but this interference forms part of a larger campaign of disruption and destabilisation which seeks to diminish trust in liberal democracies in order to weaken the West and allow for greater Russian influence around the globe [6]. By exploiting pre-existing social, political or ethnic cleavages in other countries, Russia is able to stoke further divisions among populations and undermine public societal trust, as well as trust in governance, institutions or the democratic process [4]. This degradation of trust on a societal level leads to greater polarisation and division, which allows Russia to influence susceptible voters more easily in the target country on what to think, how to think and, ultimately, how to vote [6].

As shown by the example of Russia, cognitive warfare exploits trust in order to affect society in a two-step process. First, trust must be eroded in order to weaken the fabric of society. This may be civilian trust in state institutions or trust in other institutions such as the media, science, tax authorities, or the legal system [9]. This, in turn, can affect people's trust in one another, whether that means in-group versus out-group mistrust, or even mistrust within in-groups [37]. This widespread diminishment of trust destabilises a society because a loss of trust in institutions or in one another affects the legitimacy of institutions and by consequence their capacity to function, which then undermines the social contract that binds society together [9], [37]. Where presence of trust facilitates cooperation and contributes to social cohesion, the presence of distrust does the opposite [20].

Second, once trust has been eroded, the manipulation and influence of target populations can take place. A lack of trust diminishes the foundations of democracy and reduces a society's ability "to absorb shocks, recover quickly, adapt, innovate and develop further" [22] (p.32), which in turn facilitates further erosion. As previously stated, the aim of cognitive warfare is to sow distrust within a population in order to weaken its resolve and to create a fertile environment for manipulation and influence through the weaponisation of public opinion [6]. Malevolent actors can therefore benefit from the social unrest and division that results from a deterioration of trust and the spread of distrust within society.

Trust is therefore a key target for actors seeking to destabilise, disrupt and undermine adversaries and is an "essential component of cognitive warfare in which the human is the vulnerable target" [29] (p.5). It follows then that societies with high levels of trust are more resilient to the effects of cognitive warfare while low trust societies are more vulnerable. Therefore, according to the European Centre for Excellence for Countering Hybrid Threats (Hybrid CoE), in order to safeguard societies and the democratic processes that cognitive warfare seeks to attack "we need to consider trust as the force that binds societies together" [22] (p.31).

1.1 Aim of the Paper

As discussed, intervening to bolster trust is critical to strengthening societal resilience and to reducing vulnerability to cognitive warfare threats. Trust, however, is a broad, multi-layered, dynamic construct that is difficult to grasp and to intervene upon. The suitability and potential effectiveness of an intervention may vary depending on the specific trust dynamic. Therefore, to select the optimal trust intervention for a given situation, one must understand the dynamics of trust.

Having established the relevance of trust to the mitigation of cognitive warfare threats (Section 1), we now present the aim of the paper (1.1): to promote understanding of what trust is (Section 2) and how it works

(Section 3). This includes presenting a conceptualisation of trust that can be used to characterise the state of a trust dynamic and, in doing do, elucidate the suitability of related interventions. In the following section (3: Trust), we seek to clarify the conceptual boundaries of trust, including mistrust and distrust (2.1), types of trust (2.2) and the elusive ideal: calibrated trust (2.3). In the present paper, we describe trust as the ever-fluctuating product of ongoing trust calibration (Section 3). This entails three phases (i.e., situations) in which a trust dynamic can be characterised as being: 'building' (3.1), 'managing' (3.2) and 'repairing' (3.3). The context and nature of these phases and exemplar interventions are discussed, in an effort to bring together theory and practice. Finally, implications, limitations and key takeaways are discussed (section 4). In sum, the objective of this paper is to conceptualise the trust calibration cycle to help identify the status of a trust dynamic, provide foresight into future developments and to support selection of appropriate interventions. By reimagining the trust calibration cycle (using social science literature), we hope to support the ability of policymakers to understand, identify and intervene upon trust in society in order to bolster societal resilience as a means of mitigating the effects of cognitive warfare.

2.0 TRUST

2.1 The Trust Family

Actors engaged in cognitive warfare often seek to influence a target society by either eroding trust or by sowing distrust. Although this may seem like different ways of describing the same process, recent theoretical developments have distinguished the two mechanisms. Traditionally, trust and distrust were seen as opposing points on a single spectrum, but there is a growing consensus that they are distinct concepts with unique attributes and subject to different factors of influence [46]. Research suggests that trust and distrust are orthogonal dimensions; different concepts that can coexist [46]. They exist on separate spectrums. Hence, the absence of trust should not be equated with the presence of distrust [7].

In fact, trust is better understood as a family, including the elements of trust, mistrust and distrust [7], [21], [18], [28]. Here, trust is defined as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" [30]. While trust reflects an expectation of something positive, distrust is a distinct attitude characterised by the active expectation of negative consequences (e.g., harm) associated with trusting [7], [46], [47]. Mistrust, on the other hand, is the lack of an explicit positive expectation and implies lowered perceived trustworthiness without sufficient substantiation, leaving room for the option that a trust appraisal might be based on a misjudgement, misunderstanding or misinterpretation [41]. In practical terms, mistrust can be described as a cautious attitude that motivates citizens to keep a vigilant eye on the political and social developments within their communities [28]. Although the term 'mistrust' is sometimes used interchangeably with 'distrust', they are distinct [41].

2.2 Type of Trust

The subjects of trust, mistrust and distrust can be a variety of different actors within society, creating different types of trust. These types include trust in others (social trust), trust in the government (political trust) and trust in other organisations such as the media, science, tax authorities, or the legal system (institutional trust). Social trust can be defined as the propensity of individuals to trust others in general (both known and unknown), not directed at specific people for specific purposes [23], [26], [42]. It allows us to connect with people who are different than us and facilitates interaction between different groups in society [43], which is an important prerequisite for a well-functioning democracy. There appears to be a mutually reinforcing relationship between social trust and the functioning of a democratic state; a democratic state depends on a sense of social cohesion, but a well-functioning state also seems to enhance feelings of social trust among its citizens [42]. Since processes like polarisation pose a threat to our sense of social cohesion, they can therefore erode the fabric of society. Political trust, sometimes defined together with

institutional trust, refers to citizens' trust in the political process and democratic system, as well as citizens' confidence in and support for political institutions, such as the executive and the legislature, in the face of uncertainty about or vulnerability to the actions of these institutions [31]. Institutional trust therefore refers more to other societal institutions, such as the judiciary, police, tax authorities or even media organisations.

These three types of trust are interconnected; if one type of trust is affected, it can influence the others. For example, scholars have argued that distrust and suspicion towards societal and political institutions can gradually erode social trust [37]. Institutions within society moderate citizens' social relationships both by providing people with a sense of safety and by establishing and upholding common norms and values [37]. If institutions fail to do this, it undermines the embeddedness of foundational values and principles in society and thereby risks eroding citizens' trust in society itself. Understanding these interdependencies helps us to see how trust works in different parts of our lives and how trust is the backbone of a stable society.

A cautious attitude of mistrust can help us adopt a kind of vigilance that is needed to maintain a healthy democracy [28]. Distrust on the other hand can, in the worst case, turn into political cynicism; "the attitude that assumes the worst of the nature of political objects (actors, institutions)" [31] (p.6) and the belief that "the political process and its actors are inherently corrupt, incompetent and self-serving" [32] (p.5). As such, distrust is deemed detrimental to democracy, while trust – and mistrust – are central to its success [28].

2.3 Calibrated Trust

To bolster societal resilience, the objective should be to achieve a healthy balance of trust and mistrust while also working to prevent distrust. The literature to date has tended to focus on building trust and mitigating distrust, while the role of mistrust in creating a healthy balance of trust in society has been neglected. This optimal level of trust may be referred to as 'calibrated trust'. This term originates from the concept of trust in automation literature, where "calibrating'" trust was introduced as a term to describe the process of "setting users' trust in decision aids to an appropriate level" [35]. Since then, this has become a widely known concept in human-automation literature [26], especially in relation to the usage of increasingly autonomous machines in consequential domains such as the military. In these domains, miscalibration, represented as either 'over'-trust or 'under'-trust, can lead to inappropriate reliance and can compromise safety and profitability respectively [5], [25], [45]. In other words, the concept of calibrated trust has two important qualities. First, it highlights that the amount of trust in another agent should be warranted by that agent's actual trustworthiness. Second, finding that balance demands an ongoing process of recalibration [26].

The trustworthiness of any actor varies across time and context. For example, it may be reasonable to trust an individual to be able to perform one task but not another based on their skill set (e.g., while you may trust an elected official to draft legislature, you may not trust them to make medical decisions, or vice versa). Furthermore, their trustworthiness on a given day may vary depending on their personal circumstances as well as on external situational factors (e.g., while you may trust an employee to manage the shop on a quiet weekday, you may not trust them to handle a busy weekend shift, or during a tumultuous time in their personal life). Hence calibrated trust should not be viewed as the static state of trust that underwent calibration based on cumulative experience, but as the fluctuating quality of trust that is subject to continual calibration based on ever-evolving experience. Recalibration requires persistent critical reflection. As such, the term 'calibrated' does not indicate that the trust once underwent calibration and is now calibrated; rather it refers to an enduring quality of the trust: that it is subject to calibration, continually.

Trust calibration has received scant attention in the interpersonal realm. We believe, however, that due to its above-mentioned characteristics, it provides useful perspective on interventions for trust in society. First, striving for a calibrated (as opposed to a maximised) level of trust highlights the importance of critical reflection and appreciates the cautious nature of mistrust, which is deemed an important basic condition for a resilient society that is both robust and able to adapt to changing circumstances. There is great value in the ability to accurately assess a source, person or institution's trustworthiness, thereby minimising 'over'-trust

and 'under'-trust. There is also value in the democratic behaviours associated with political mistrust (e.g., protests and demonstrations), which can themselves induce behaviour change from political representatives and promote democratic values [31]. Indeed, representative democracy depends on its citizens to uphold democratic principles and values (e.g., by being willing to monitor and hold officeholders to account and ensure that trust in government is not unconditional) [31]. That said, the consequences of excessive mistrust (i.e., mistrust that exceeds what is merited by the actual trustworthiness of the trustee) are significantly different in AI than in the interpersonal domain. In the human-technology interaction domain, the consequence of excessive mistrust simply results in the trustor declining to use an optional technology and potentially missing out on its benefits (e.g., increased efficiency). In society, where trust is indispensable to the functioning of democracy, excessive mistrust can have potentially fatal consequences [9], especially if institutions and political officeholders are unresponsive to that mistrust.

Secondly, conceptualising trust calibration as an ongoing and iterative process also provides a useful perspective for intervention selection and execution. This temporal perspective of trust calibration has been priorly used in the interpersonal domain in a model that describes how people continually (re)calibrate their levels of trust and distrust in reaction to external events [1]. We believe that well-calibrated trust within society must be achieved through an ongoing process of evaluation and adjustment in order to stay attuned to changing circumstances. Consistently maintaining trust is as important as initially building it.

3.0 THE TRUST CALIBRATION CYCLE

To navigate the constant pursuit of an optimal level of trust in ever-changing circumstances, researchers in autonomous technology literature have distinguished between three stages of trust: formation, violation and repair [44]. Adapting these terms to the interpersonal domain, we propose a categorisation of calibrated trust which comprises three broad phases: building, managing and repairing trust. These phases can occur in any order and can be repeated. Given the complex nature of trust, these three phases act as labels that can be used to identify the status of a particular trust dynamic, to understand the potential next developments and recognise relevant interventions.

3.1 BUILDING

3.1.1 Process

Trust can be defined as a willingness to be vulnerable to another's actions based on the expectation that the other will act in a certain way [30]. This introduces two concepts that are key in the building phase: expectation and vulnerability. First, people assume intuitively that past behaviour is the best predictor of future behaviour. It follows that demonstrations of trustworthy behaviour would build expectations of future trustworthy behaviour and thus build (propensity to) trust. Second, vulnerability and trust are closely related. The more vulnerable one is, the greater the consequences of being let down by the trustee and the greater the weight of the trust decision. In other words, trust is a commodity. It follows that reducing one's vulnerability would build their 'trust capital', affording a greater propensity to trust.

When forming a trust appraisal, the trustor decides to either trust or not trust based on their baseline 'summary judgement' of the trustee [1]. This initial stance is the sum of many judgements of trustworthiness, which "may not be wholly internally consistent and may contain both elements of trust as well as distrust" [1] (p.3). Building trust is crucial to creating a positive (i.e., trusting) initial summary judgment. It is also crucial for providing appropriate counterweight to mistrusting instincts, thereby preventing excessive mistrust. In summary, consistently working to build and bank up trust across types and time is paramount.

3.1.2 Intervention

Building trust can involve a raft of interventions implemented by a wide range of actors across the entire breadth of society. Interventions aimed at building trust tend to be more sustained, both in duration of execution and in realisation of effects. Building and establishing trust – whether trust in institutions, within society, or between particular groups or individuals – is a mammoth endeavour and an undoubtedly daunting task for any prospective practitioners charged with undertaking it. However, there are some long-term interventions that can be implemented to build trust within society across different types of trust.

Trust-bolstering interventions often revolve around the theory of **reciprocity**: that bestowing trust engenders trust [38, 16]. A recent study put this to the test and analysed the effects of **alternative social welfare policies** (inspired by universal basic income), which were based on a more trusting and unconditional approach than traditional welfare policies. The authors' hypotheses were based on the fundamental argument of reciprocity: "a trusting government will harvest trust from welfare recipients in return" [8]. The main finding was that the more trusting policy with less government control and more autonomy increased local trust among recipients [8]. The link between social welfare policies and trust is supported by other literature that suggests reducing social inequality and depravity can increase societal trust [48], this provides strong evidence that a causal relationship exists between the bestowing of trust upon individuals and the receipt of reciprocated trust [8].

To benefit from trust reciprocity requires significant investment and long-term strategising. A shorter-term and perhaps more tangible trust-bolstering intervention is to **increase political participation** which, as research suggests, can increase political and social trust, especially among marginalised groups [11], [23], [47], [24]. Based on this premise, Weymouth et al. [47] put forth '**deliberative democracy mini-publics**' as an example of an effective intervention to help build political trust. A mini-public involves a representative group of citizens (i.e., citizens randomly selected from the population) who gather to engage in small-group discussions on a topic of importance. The discussions are assisted by an independent facilitator, with the aim of reaching collective decisions or recommendations. Based on two use cases, the authors validated the effectiveness of this tool as a means to increase participation and political trust [47]. Participation, or the availability and ability to participate, in the political process is therefore key to societal resilience building. Other research has found that combining participatory programs with broad informational campaigns can increase participation and social cohesion, as well as political and societal [24].

Similar to political participation is **social inclusion**. As discussed earlier, a main objective in cognitive warfare is to enflame pre-existing divisions in a society in order to create greater polarisation and fragmentation which can be exploited for political or military gain. It follows, then, that in order to strengthen societal resilience against the threat of cognitive warfare, there must be a concerted effort to reduce division and polarisation by promoting social inclusion and integration. As such, Wigell et al. [48] recommend implementing targeted programmes aimed at integrating diasporas through greater awareness, closer social proximity and the promotion of the role of civil society actors from and within marginalised groups. A greater diversity of civil society actors and of representation results in higher levels of trust [14].

3.2 Managing

3.2.1 Process

While the building phase aims to grow trust, the managing phase seeks to maintain prior or pre-existing levels of trust. Despite best intentions, trust violations do occur, meaning damage control is needed to protect pre-existing trust. When the actual outcome of a trust decision does not match the expected outcome (i.e., after a violation of trust), recalibration is initiated wherein the trustor evaluates what went wrong, why and what they should do differently in the future. Adams [1] outlines the steps of this evaluation in their model of the trust calibration cycle. After the violation, the trustor first engages in sense-making. This involves

assessment of responsibility either to the self or to the trustee. If the trustor deems themselves responsible, they may proceed to re-evaluate their approach to judging the trustworthiness of others. If the trustee is deemed responsible, assessment of intentionality follows. If the intention is deemed not to have been malevolent, trust levels may decrease towards mistrust. If, however, the intention is deemed to have been malevolent, this sets the stage for active distrust to emerge.[1] Although this model portrays a simplistic example (a discrete decision situation with a unidimensional summary judgment, an explicit trust decision outcome and binary responsibility and intention attribution), it nevertheless illuminates opportunities for intervention within the trust management phase (e.g., by influencing the responsibility and intentionality assessments). Trust management may look like damage mitigation after a violation or, in the absence of a violation, like the maintenance of existing trust through reliable action.

3.2.2 Intervention

When dealing with a breach or violation of trust, a fine balancing act is required so that the damage caused by the violation is managed and mitigated to avoid further denigration of the trust dynamic. The response therefore is inherently more reactive than interventions required in the building phase: a violation has occurred which needs to be addressed specifically, directly and immediately to prevent further spread of distrust or halt a dive in trust. Interventions aimed at managing the effects of a trust violation must be considerate of those whose trust has been violated. Reticence is a suboptimal response to a perceived trust violation [17]. When people start to doubt the intentions or trustworthiness of an actor and start to suspect malevolent intent, it can set the stage for distrust [1], which is wholly undesirable.

Therefore, **open, transparent and competent communications** are key to managing trust violations. Research on risk and crisis communication often points to features such as demonstrating honesty, empathy, competency [33], openness and shared interest [15]. One of the most important factors in the production, maintenance and restoration of trust, especially in times of crisis or scandal, is transparency [13]. Transparency is effective because it helps build confidence via the process of accountability and fairness [13]. Two types of transparency have been identified in the literature: (1) reputation of transparency; and (2) efforts to communicate transparently [3]. Demonstrating both types of transparency achieves a greater level of trust than demonstrating either alone or neither [3], meaning that to effectively manage trust, transparency should be consistent, if not habitual. Transparency alone is not enough. It has far greater power when it is combined with the other qualities of effective communication mentioned previously, such as competent performance, empathy, benevolence and openness [13]. Good communication also allows a violation to become an opportunity for a trustor to learn about the trustee's true capabilities and to hone their mental model of the trustee's trustworthiness accordingly (i.e., to better calibrate their trust).

Context is also important to effective communication surrounding trust violations. Research has found that private organisations achieve greater levels of trust when demonstrating openness, while public organisations garner trust most significantly when they demonstrate competency [40]. Furthermore, trust also depends on the balance between perceived competency and perceived motives. That is to say, if an actor is perceived as being motivated to withhold or distort information then public trust will be undermined regardless of how knowledgeable or expert the actor is perceived to be [15]. Therefore, the principles of transparency, openness and competency complement one another and should be used in combination when trying to effectively manage trust in response to a trust violation.

Another reactive intervention that can mitigate lowering trust or growing distrust in the event of a trust violation is **debunking**. This is perhaps only relevant in cases of cognitive warfare strategies whereby the trust violation occurs because of the spread of mis- or disinformation. In this situation, a common strategy to manage levels of trust and distrust is to refute the violating information and to provide evidence of its falsity.

[1] The recalibration of trust also takes place when a violation occurs in opposition to the initial trust appraisal. In other words, when an initial decision not to trust is violated by an unexpectedly positive action by the trustee, this may lead to a reduction in distrust [1].

However, responding to trust violations such as these is "not merely a technical problem" and requires a response beyond mere debunking of the information [4]. Critical thinking and a healthy scepticism of information should be nurtured within a population [13]. This is often referred to as **pre-bunking** as it relates more to pre-emptive preventative measures rather than the more reactive measures of debunking. Pre-bunking can comprise several longer-term measures which are designed to stimulate critical thinking and engender trust empowerment. These include an increase in digital literacy, media literacy and greater education on the threats of hybrid threats and cognitive warfare, particularly mis- and disinformation. We consider pre-bunking a relevant management intervention because it can be employed concurrently with a disinformation campaign. For instance, actors can design and communicate public information campaigns on the relationship between disinformation and societal trust that supplements a wider drive to improve education on media literacy and disinformation [48].

These interventions are most applicable after a trust violation has taken place. However, pre-bunking, for example, does not necessarily need to be implemented in response to a trust violation but can also be implemented on a continual basis to uphold and maintain trust levels. It is good practice to sustain these interventions for the ongoing maintenance of trust, whether a trust violation occurs or not.

3.1 Repairing

3.3.1 Process

The repair phase focusses on restoring trust that has been damaged [36]. It follows violations that were not (entirely) successfully managed, i.e., that led to reduced trust, increased mistrust, or even sowed seeds of distrust. Context is what separates the building phase from the repair phase; it is the difference between building and rebuilding. While trust can be broken swiftly, repairing trust is a process that takes time, patience and commitment. This is because building trust is harder when starting from a position of distrust (i.e., rebuilding) [36]. Effects will therefore not be direct or automatic, so a long-term perspective is needed, but repair is possible [36].

3.3.2 Intervention

To re-establish trust in an environment of mis- and distrust, it is crucial to understand the damage that caused these latter two, as this can inform better strategies for rebuilding the trust. A large psychological literature review by O'Brien & Tyler [36] proposed two evidence-informed trust-bolstering strategies that authorities can employ together to rebuild community trust: reconciliation and the procedural justice approach.

Reconciliation entails actively listening to the concerns of those who have experienced damaged trust and making community-level trust-fortifying gestures that address both the past and the future. These gestures should both demonstrate recognition of the past injustice and its harm and promise changed future relations with or behaviour by the authority. They should also include responsibility acceptance and/or an apology [36].

Following **the procedural justice approach**, legal and political authorities can re-earn the public's trust and counteract negative beliefs by altering their behaviour towards the public and the way they exercise authority [36]. Authorities should demonstrate commitment to their promises of positive change by putting persistent effort in delivering concrete results that benefit the community or stakeholders. Procedural justice strategies focus on treating members of the community fairly and respectfully in daily interactions.

O'Brien and Tyler highlight the importance of deploying both strategies simultaneously. Reconciliatory gestures are not a substitute for change in daily practices. The authors emphasise the role of process in reconciliation; trust is not (re)built in a single event, but rather can occur when concrete procedural justice improvements are accompanied by a series of gestures. The success of both strategies stands or falls with to the extent to which they are perceived as sincere [36]. Furthermore, the implementation of one strategy can reinforce or, inversely, undermine the perceived sincerity of the other [36].

4.0 DISCUSSION

4.1 Key Takeaways

4.1.1 Trust Calibration is a Continuous Process

We have highlighted that calibrated trust should not be viewed as a single static state attributed to multiple targets. An individual does not (and should not) have one level of calibrated trust, because the appropriate level of trust varies constantly depending on who or what it is attributed to, when and in what context. The trustworthiness of a trustee differs across time and task. All in all, finding an appropriate balance between trust and mistrust requires persistent critical reflection on a range of factors. Searching for that optimised state between 'over'-trust and 'under'-trust is crucial for a resilient society that is both robust and able to adapt to changing circumstances. Recalibration is an ongoing process based on ever-evolving experience.

It is important to recognise that our categorisation should be viewed a simplification of reality, as it distils the complex trust process into phases which are often less distinct in practical application than they appear in our trust cycle. As mentioned, the sequence of these phases is flexible and there is often overlap and blending of concepts and processes. Hence it is crucial to consider the nuances and to interpret our categorisations within the broader context of the multifaceted, dynamic nature of trust to gain a more accurate understanding of the interactions at play.

4.1.2 Bolstering Trust Requires Psychosocial Interventions

In our proposed categorisation of trust, we have drawn from different academic disciplines to suggest a range of psychosocial interventions to influence trust. This sample of interventions, though by no means exhaustive, illustrates the insight that stands to be gained from the social sciences. Cognitive warfare aims to control how people react to information and to change how a target population thinks and, in turn, acts [6], [4], [10]. Since the goal of malevolent actors is to interfere with the *minds* of their opponents, interventions to counter these threats should likely come from the psycho-social domain as well. "Responding to a cognitive warfare strategy is not merely a technical problem.*"* [4] (p. VI). Hence, technical solutions such as improving the security of digital voting systems will not be sufficient to prevent the Kremlin from undermining Western elections by interfering with the minds of voters [4]. Instead, solutions to cognitive warfare threats should emerge from the human and social sciences, draw on human factors methodology and engineering and be based on models of the cognitive processes in question [12].

It must be noted that although we stress the importance of critical reflection, trust is something that happens mostly unconsciously and is largely driven by emotions and 'gut-feeling'. Indeed, disinformation narratives are often targeted to trigger emotions of fear, anger, disgust, confusion and hopelessness, with the aim to create suspicion towards and distrust in governmental institutions [19]. Although we emphasise critical reflection and rational thought throughout our paper, the interventions discussed also address emotion, albeit more indirectly. For example, political engagement and social cohesion interventions would likely influence feelings of inclusion and isolation, reconciliation interventions would impact feelings of hurt and procedural justice interventions may create a sense of safety. The notion that emotion and 'gut-feeling' are important drivers of trust only underscores the need to involve social scientists (for whom emotion is an area of expertise) in the mission to strengthen societal resilience through intervention on trust. As previously stated, this is a people problem, not a technical problem. As such, it requires a people-focused solution.

4.1.3 Building Resilience Via Trust Requires a 'Whole-of-Society' Approach

When faced with cognitive warfare, trust as a form of societal resilience is key. However, as the above exploration of trust underlines, no single intervention alone can establish a sufficiently resilient level of trust

in society. Bolstering trust as a means of countering the complex and interrelated threats associated with cognitive warfare therefore requires a "well-functioning whole-of-society approach" [48]. This means that policymakers should adopt an approach to trust-bolstering that "integrates the capacities and capital possessed by various private and civil society sector" into one collaborative, unified effort [48]. Therefore, to achieve a sufficiently healthy level of trust calibration requires the implementation of a multitude of interventions implemented incrementally and on an ongoing basis across all levels of society; from micro to macro and top-down to bottom-up. Furthermore, this must involve the collaboration of a range of actors - from civil society and the public sector to private companies and individuals - to engage in proactive resilience building against the complex threats posed by cognitive warfare [39]. Of course, successful collaboration of key actors within a 'whole-of-society' may be dependent on individual, group or cultural differences that exist within society. For instance, one group may give more weight to the values of integrity than to competence or value communalism more than individualism. Such differences should be borne in mind when devising resilience-building interventions.

The notion of a 'whole-of-society' approach to resilience-building is not novel and the concept is becoming increasingly central to policymaking in the European Union around hybrid threats, such as cognitive warfare. In fact, a report by the Hybrid CoE in 2023 advocates for a 'whole of society' approach to building resilience against hybrid threats and sets forth a comprehensive resilience ecosystem model designed to achieve this [22]. Further still, the report highlights trust as one of foundations of the democratic system and should be at the heart of societal resilience-building [22]. This largely reflects the findings of this paper; however it is argued here that such a simple conceptualisation of trust is somewhat unhelpful and that a more nuanced understanding of trust is required to fully appreciate the depth and breadth of response and interventions needed to bolster trust as a form of societal resilience.

As well as 'whole-of-society', building, maintaining and repairing trust in the face of cognitive warfare threats also requires a 'whole-of-trust' approach. There are various different types of trust (with political, societal and institutional being the focus of this paper) and various branches of the trust family (trust, mistrust and distrust). The interconnectedness of these different types of trust means that changes to the level of one type of trust or distrust can have consequences for another. For instance, if a trust violation reduces political trust, then it is likely that institutional trust, particularly in legislative or judicial institutions, will also decline. Building trust, therefore, requires a comprehensive approach that ensures all types of trust remain calibrated to a sufficient degree. In addition to the different types of trust, there are different phases of the trust life cycle, as discussed above, which also need to be taken into account. A 'whole-of-society' and 'whole-of-trust' approach to bolstering trust therefore requires a fully comprehensive engagement with trust, both in terms of the type of trust and where in the cycle of trust the intervention takes place.

4.2 Conclusion

This paper argues that trust must become "the key bulwark" in deterring and resisting cognitive warfare and bolstering trust should be the "linchpin of efforts to neutralise hybrid warfare" [9]. However, current conceptualisations of trust are limited and a more nuanced understanding of trust is required to fully appreciate the depth and breadth of response and interventions needed to bolster trust as a form of societal resilience. The strength of trust relies not only upon relations between the state and the individual but also upon relationships between individuals and the social cohesion and mutual trust that that fosters [2], [6]. As Bernal et al. [6] posited:

> *The foundation for democracy lies not only in laws and civil order, but also in trust and mutual respect: the trust that we will follow those laws, respect civil institutions and respect each other and our differing opinions.* (p.4).

Therefore, in order to counter the complex weave of threats associated with cognitive warfare, a 'whole-of-society' and 'whole-of-trust' approach should form the basis of trust-based interventions, an approach which

encompasses all levels, from state and societal to individual and interpersonal – it affects and is affected by all levels in society. This is why, as Bilal [9] concludes, any solutions against cognitive warfare that do not incorporate the fortification of trust in society "will probably fall short of offering effective antidotes" and "nothing will work or produce the desired results in the absence of trust". It is therefore imperative for policymakers and strategists to put trust at the heart of resilience building.

5.0 REFERENCES

[1] B. D. Adams, "Calibration of Trust and Distrust," Toronto, CA-ON, 2005.

[2] A. C. Apostol, N. Cristache, and M. Năstase, "Societal Resilience, A Key Factor in Combating Hybrid Threats," Int. Conf. Knowledge-Based Organ., vol. 28, no. 2, pp. 107–115, Jun. 2022, doi: 10.2478/KBO-2022-0057.

[3] G. A. Auger, "Trust Me, Trust Me Not: An Experimental Analysis of the Effect of Transparency on Organizations," J. Public Relations Res., vol. 26, no. 4, pp. 325–343, 2014, doi: 10.1080/1062726X.2014.908722.

[4] O. Backes and A. Swab, "Cognitive Warfare- The Russian Threat to Election Integrity in the Baltic States," Belfer Cent. Sci. Int. Aff., no. November, 2019, [Online]. Available: www.belfercenter.org.

[5] A. L. Baker, K. E. Schaefer, and S. G. Hill, Teamwork and Communication Methods and Metrics for Human – Autonomy Teaming. 2019.

[6] A. Bernal, C. Carter, I. Singh, K. Cao, and O. Madreperla, "Cognitive Warfare: An Attack on Truth and Thought," 2020. doi: 10.31826/jlr-2015-120101.

[7] E. Bertsou, "Rethinking political distrust," Eur. Polit. Sci. Rev., vol. 11, no. 2, pp. 213–230, 2019, doi: 10.1017/S1755773919000080.

[8] J. Betkó, N. Spierings, M. Gesthuizen, and P. Scheepers, "How Welfare Policies Can Change Trust – A Social Experiment Assessing the Impact of Social Assistance Policy on Political and Social Trust," Basic Income Stud., vol. 17, no. 2, pp. 155–187, 2022, doi: 10.1515/bis-2021-0029.

[9] A. Bilal, "Hybrid Warfare – New Threats, Complexity, and 'Trust' as the Antidote," NATO Review, 2021.

[10] K. Cao, S. Glaister, A. Pena, D. Rhee, R. William, and A. Rovalino, "Countering Cognitive Warfare—Awareness & Resilience," NATO Review, 2021. https://www.nato.int/docu/review/articles/2021/05/20/countering-cognitive-warfareawareness-and-resilience/index.html (accessed Feb. 03, 2023).

[11] T. Christensen and P. Lægreid, "Trust in Government: The Relative Importance of Service Satisfaction, Political Factors, and Demography," Public Perform. Manag. Rev., vol. 28, no. 4, pp. 487–511, 2005.

[12] B. Claverie, B. Prébot, N. Buchler, and F. Du Cluzel, "What is Cognition? And How to Make it One of the Ways of the War," in Cognitive Warfare: The Future of Cognitive Dominance, NATO, 2021.

[13] A. Cole, J. S. Baker, and D. Stivas, "Trust, Transparency and Transnational Lessons from COVID-19," J. Risk Financ. Manag. 2021, Vol. 14, Page 607, vol. 14, no. 12, p. 607, Dec. 2021, doi: 10.3390/JRFM14120607.

[14] F. Ehrke, S. Bruckmüller, and M. C. Steffens, "A double-edged sword: How social diversity affects trust in representatives via perceived competence and warmth," Eur. J. Soc. Psychol., vol. 50, no. 7, pp. 1540–1554, 2020, doi: 10.1002/ejsp.2709.

[15] J. R. Eiser, T. Stafford, J. Henneberry, and P. Catney, "'Trust me, I'm a Scientist (Not a Developer)': Perceived Expertise and Motives as Predictors of Trust in Assessment of Risk from Contaminated Land," Risk Anal., vol. 29, no. 2, pp. 288–297, Feb. 2009, doi: 10.1111/J.1539-6924.2008.01131.X.

[16] E. Fehr and S. Gächter, "Fairness and Retaliation: The Economics of Reciprocity," J. Econ. Perspect., vol. 14, no. 3, pp. 159–181, 2000, doi: 10.1257/JEP.14.3.159.

[17] D. L. Ferrin, P. H. Kim, C. D. Cooper, and K. T. Dirks, "Silence Speaks Volumes: The Effectiveness of Reticence in Comparison to Apology and Denial for Responding to Integrity- and Competence-Based Trust Violations," J. Appl. Psychol., vol. 92, no. 4, pp. 893–908, 2007, doi: 10.1037/0021-9010.92.4.893.

[18] R. Hardin, Trust and Trustworthiness. Russell Sage Foundation, 2002.

[19] A. Hoyle, T. Powell, B. Cadet, and J. Van De Kuijt, "Influence pathways: Mapping the narratives and psychological effects of Russian COVID-19 disinformation," in Proceedings of the 2021 IEEE International Conference on Cyber Security and Resilience, CSR 2021, Jul. 2021, pp. 384–389, doi: 10.1109/CSR51186.2021.9527953.

[20] M. P. Jasinski, "Social trust and its origins," in M.P. Jasinski (Ed.), New York: Palgrave Macmillan, 2011, pp. 47–62.

[21] W. Jennings, G. Stoker, V. Valgarðsson, D. Devine, and J. Gaskell, "How trust, mistrust and distrust shape the governance of the COVID-19 crisis," J. Eur. Public Policy, vol. 28, no. 8, pp. 1174–1196, 2021, doi: 10.1080/13501763.2021.1942151.

[22] Jungwirth R. et al., "Hybrid threats: a comprehensive resilience ecosystem," Publications Office of the European Union, Luxembourg, 2023. doi: 10.2760/37899.

[23] M. Karlsson, J. Åström, and M. Adenskog, "Democratic Innovation in Times of Crisis: Exploring Changes in Social and Political Trust," Policy and Internet, vol. 13, no. 1, pp. 113–133, Mar. 2021, doi: 10.1002/POI3.248.

[24] D. M. Kiwanuka, "Building Trust and Reciprocity through Citizen Participation and Transparency: Lessons from Municipal Governments of Uganda and Thailand," Int. J. Econ. Bus. Manag. Res., vol. 06, no. 05, pp. 50–63, 2022, doi: 10.51505/ijebmr.2022.6505.

[25] E. S. Kox, L. B. Siegling, and J. Kerstholt, "Trust development in military and civilian Human-Agent Teams: the effect of social-cognitive recovery strategies," Int. J. Soc. Robot., 2022, doi: 10.1007/s12369-022-00871-4.

[26] J. D. Lee and K. A. See, "Trust in Automation : Designing for Appropriate Reliance," vol. 46, no. 1, pp. 50–80, 2004.

[27] J. Lee, "Post-disaster trust in Japan: the social impact of the experiences and perceived risks of natural hazards," 2019, doi: 10.1080/17477891.2019.1664380.

[28] P. T. Lenard, "Trust Your Vompatriots, but Count Your Change: The Roles of Trust, Mistrust and Distrust in Democracy," Polit. Stud., vol. 56, no. 2, pp. 312–332, 2008, doi: 10.1111/j.1467-9248.2007.00693.x.

[29] J. M. Masakowski, Y. R., & Blatny, "Mitigating and Responding to Cognitive Warfare," 2023.

[30] R. C. Mayer, J. H. Davis, and D. F. Schoorman, "An Integrative Model of Organizational Trust," Acad. Manag. Rev., vol. 20, no. 3, pp. 709–734, 1995, doi: 10.1109/GLOCOM.2017.8254064.

[31] T. W. G. van der Meer, "Political Trust and the 'Crisis of Democracy,'" Oxford Res. Encycl. Polit., no. January 2017, pp. 1–23, 2017, doi: 10.1093/acrefore/9780190228637.013.77.

[32] T. W. G. van der Meer and S. Zmerli, "The deeply rooted concern with political trust," in Handbook on Political Trust, 2017, pp. 1–15.

[33] L. S. Meredith, D. P. Eisenman, H. Rhodes, G. Ryan, and A. Long, "Trust Influences Response to Public Health Messages during a Bioterrorist Event," J. Health Commun., vol. 12, no. 3, pp. 217–232, 2007, doi: 10.1080/10810730701265978.

[34] R. S. Mueller, "The Mueller Report: Report On The Investigation Into Russian Interference In The 2016 Presidential Election," Washington, D.C., 2019. [Online]. Available: https://www.justice.gov/storage/report.pdf

[35] B. M. Muir, "Trust between humans and machines, and the design of decision aids.," Int. J. Man. Mach. Stud., vol. 27, no. 5–6, pp. 527–539, 1987.

[36] T. C. O'Brien and T. R. Tyler, "Rebuilding trust between police & communities through procedural justice & reconciliation," Behav. Sci. Policy, vol. 5, no. 1, pp. 35–50, 2019, doi: 10.1353/bsp.2019.0003.

[37] J. W. van Prooijen, G. Spadaro, and H. Wang, "Suspicion of institutions: How distrust and conspiracy theories deteriorate social relationships," Current Opinion in Psychology, vol. 43. Elsevier Ltd, pp. 65–69, 2022, doi: 10.1016/j.copsyc.2021.06.013.

[38] B. Rothstein and E. M. Uslaner, "All for One: Equality, Corruption, and Social Trust," World Polit., vol. 58, no. 1, pp. 41–72, Oct. 2005, doi: 10.1353/WP.2006.0022.

[39] D. Snetselaar, G. Frerks, L. Gould, S. Rietjens, and T. Sweijs, NATO Review "Knowledge security: insights for NATO," 2022.

[40] T. Strand Offerdal, S. Nørholm Just, and Ø. Ihlen, "Public Ethos in the Pandemic Rhetorical Situation: Strategies for Building Trust in Authorities' Risk Communication," 2021, doi: 10.3316/INFORMIT.099047207644909.

[41] J. Tully, "Trust, Mistrust and Distrust in Diverse Societies," in Trust and Distrust in Political Theory and Practice: The Case of Diverse Societies, D. Karmis and F. Rocher, Eds. McGill-Queen's University Press, 2019, pp. 1–59.

[42] E. M. Uslaner, The Oxford Handbook of Social And Political Trust. New York: Oxford University Press, 2018.

[43] E. M. Uslaner, "Does Diversity Drive Down Trust?," Nota di Lav., no. 69, 2006.

[44] E. J. de Visser et al., "Almost human: Anthropomorphism increases trust resilience in cognitive agents.," J. Exp. Psychol. Appl., vol. 22, no. 3, pp. 331–349, Sep. 2016, doi: 10.1037/xap0000092.

[45] E. J. de Visser et al., "Towards a Theory of Longitudinal Trust Calibration in Human–Robot Teams," Int. J. Soc. Robot., pp. 1–20, Nov. 2019, doi: 10.1007/s12369-019-00596-x.

[46] S. Van De Walle and F. Six, "Trust and distrust as distinct concepts: Why studying distrust in institutions is important," J. Comp. Policy Anal., vol. 16, no. 2, pp. 158–174, 2013, [Online]. Available: http://www.tandfonline.com/action/journalInformation?journalCode=fcpa20

[47] R. Weymouth, J. Hartz-Karp, and D. Marinova, "Repairing political trust for practical sustainability," Sustain., vol. 12, no. 17, pp. 1–25, 2020, doi: 10.3390/su12177055.

[48] M. Wigell, H. Mikkola, and T. Juntunen, "Best practices in the whole-of-society approach in countering hybrid threats, no. May." Brussels: European Union, 2021.