THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Declining Trust in Institutions:

Considerations for a Behavioural Intervention

ANNABEL GILLARD

KANGAN MALIK

Table of Contents

1 Introduction

The context for this paper is the widespread belief that public trust in institutions is declining, and we considered prospects for a behavioural intervention to reverse this trend. First we sought to validate the perception that institutional trust is in decline by analysing three trust surveys produced by the UN, the OECD, and Edelman, but found a mixed and incomplete picture. The results of these surveys are in conflict: both among self-reported surveys, and when comparing reports to behaviour. Secondly we analysed different models of trust to identify its components and drivers and found that ethics is an important feature of earned trust, along with competence and honesty. We note that self-reported surveys identify only the public perception of trust but misses whether the institutions deserve it. Therefore, surveys only cover one side of the trust relation. The environment in which the trust relation is established also has an effect - but published data on the interactions between environment and trust perception is inconsistent. We conclude that enacting behavioural interventions to increase trust perception would only be ethical if it were to correct an inaccuracy. Without an accurate assessment of whether trust is justified by an institution's underlying trustworthiness we contend that such an intervention would be unethical. It would not only undermine the case for trust in that institution, but also potentially trigger moral injury. We suggest some areas for future investigation to inform the most appropriate, effective and ethical intervention.

2 Context and Purpose

2.1 Background

In this white paper, we sought to better understand the drivers of declining trust in institutions and whether behavioural science could have a role in addressing this problem. The motivation for this research is that collective action on urgent political, social or economic questions is typically coordinated through institutions. Trust in those institutions confers legitimacy and motivates public engagement with their recommendations (OECD[1]). Most pressingly, addressing the climate crisis will likely require sacrifices from citizens, such as changing consumption habits or accepting a lower level of economic activity. The Covid-19 pandemic demonstrated that public acceptance and compliance with policies that require personal sacrifice, such as lockdowns, are greater where institutions are trusted (LSE[2]). By implication, where the success of public policy is a life-or-death matter, increasing

[1] https://www.oecd.org/governance/trust-in-government/

[2] https://blogs.lse.ac.uk/impactofsocialsciences/2021/05/06/is-a-breakdown-in-trust-transparency-and-social-cohesion-a-price-worth-paying-for-more-extensive-data-linkage/

trust in institutions will enable greater public support and adoption of policy with material public benefit.

We assessed the leading trust surveys provided by the UN, the OECD and Edelman, and found that they do not show a consistent picture. The UN and OECD show a declining trend of trust in government, but in contrast the Edelman survey shows a gradual rise over the past decade, albeit at low absolute levels. Further analysis on Edelman shows that these averages mask a number of underlying dynamics. These contradictions raise the question of how to reliably assess perceptions of trust. We analyse different methodologies and compare them to each other and to key concepts from literature review.

The second question we wish to address is the relation of trust perception to underlying trustworthiness. We believe this question needs to be resolved prior to any intervention for two reasons: ethics and efficacy. Interventions are most effective when targeted and contextual, which requires an understanding of the drivers of the trust shortfall (BMJ[3]). If declining trust accurately reflects the trustworthiness of the institution, then an intervention to falsely inflate perceptions of trust appears unethical. Such interventions also risk undermining trust in behavioural science and amplifying mistrust in institutions. We analyse the relatively understudied field of measuring trustworthiness and conclude by outlining possible areas for future investigation.

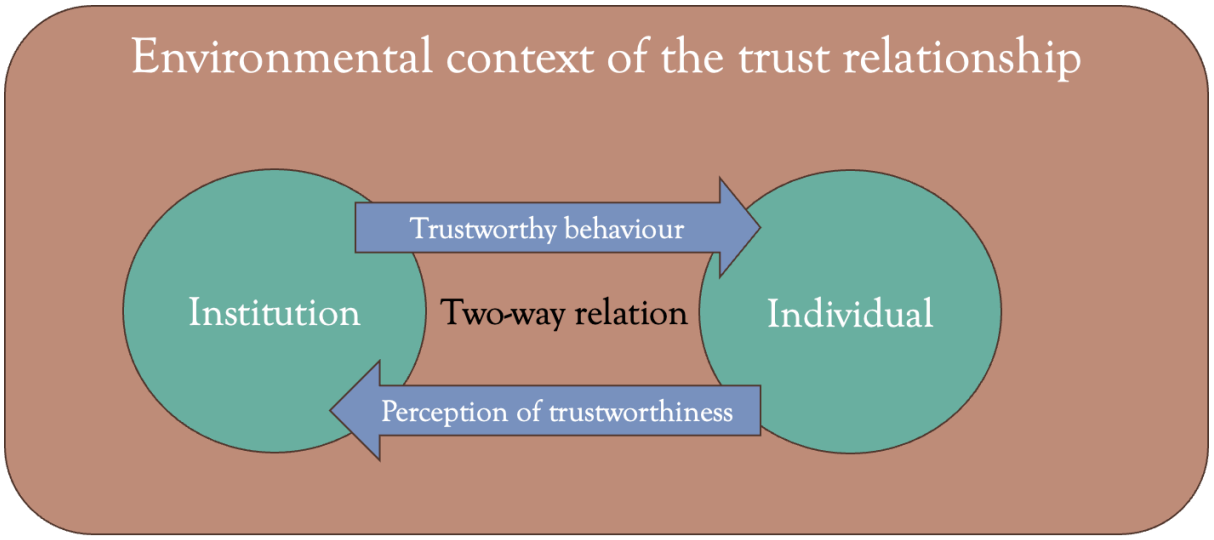
2.2 What is Trust?

Although the literature has not cohered around a single consensus definition of trust, certain components feature repeatedly, particularly reliability, honesty or integrity, and competence (Ashraf et al, 2016; O'Neill, 2018; TiGRE report, 2020). A recent study on trust in the corporate sector identified four elements of trust: competence, motive, means and impact (Sucher and Gupta, 2022). In practical terms, ***can*** the institution perform its task (competence), ***will*** the institution perform its task (reliability/ impact) and do the actions match the words (integrity/ motive). These elements also apply to government sector institutes as per the OECD[4] Trust in Public Institutions framework.

Figure 1 illustrates the philosophical perspective in which trust is defined as a *relation* between two parties, situated within a *climate* (MacLeod, 2020). It cannot be forced, but must be cultivated.

[3] https://jech.bmj.com/content/73/6/481
[4] https://www.oecd.org/governance/trust-in-government/

Figure 1: Illustration of components of trust relationship between individual and institution



Trust can also be classified as outcome-based, or process based. When people's trust in their governments is based on performance and expectations, this is known as "outcome-based" trust and is a direct factor in the government's performance and individual's expectations. However, process-based trust is determined by people's engagement experience with the government and is a factor of citizens' satisfaction with the level, depth, and quality of engagement (e.g., type, and frequency) in the government's decision-making process (World Bank, 2020).

The observation: *"Trust is important, but it is also dangerous"* (Macleod, 2020) summarises why trust matters - it is important because of the instrumental value it offers, and it is dangerous because it involves risk. Trust differs from knowledge in that trust relates to the unknown - it is the confidence that the unknown will meet expectations. Trust is conditional, built on the components outlined above such as competency, integrity and reliability etc. The vulnerability this confers is such that when trust is betrayed, it is traumatising, described as 'moral injury' (Shay 2014). Most of the studies into moral injury have focussed on childhood trauma or experiences in front-line military and health services and identify that a transgression of one's values or a breach of trust can not only result in immediate emotional harm but also has the lasting effect of reducing the capacity to trust in future (Shay 2014; Brennan, 2021; Kidwell and Kerig 2021; Williamson et al 2021). Thus, those responsible for upholding trustworthy institutions need to be mindful not only of their role in the collective benefits of conferring political legitimacy and supporting economic prosperity but also of their moral obligation and duty of care to individuals.

2.3 Why Trust Matters?

Trust is a key element in numerous academic disciplines, underlining its importance to human relations and understanding. Philosophers debate whether it has intrinsic value, but agree that it has clear instrumental value and is an integral element of ethical discourse (MacLeod, 2020). The economic and political benefits are also well documented, and the importance of stable and trusted institutions is demonstrated by their representation in the UN's Sustainable Development Goals (SDGs[5]).

Politically, the consent of the population to pay taxes, submit to laws and accept the authority of government, known as the social contract, depends on public confidence in institutions (Murtin et al, 2018). Economic benefits operate at the individual, corporate and institutional levels with higher trust translating into greater willingness to invest, take risks and navigate disruption (Barney and Hansen, 1994; Chang, Dillon, and Hussain, 2006). Moreover, the erosion of trust in government is thought to indicate "the crisis of democracy" with direct and severe consequences for the quality and ability of representative democracy, its institutions, and its actors (Crozier, Huntington, and Watanuki 1978; van der Meer 2017).

In addition to material and political benefits, there are psychological benefits. Research suggests that individuals value cooperation and have a preference for altruism and reciprocity, which reduces as trust declines in institutions, leading to instability (Bowles & Polanía-Reyes, 2013). Moreover, research shows that public trust in the government is negatively correlated to the psychological distress the public bears (Olagoke et al., 2020). Overall, the psychological and emotional benefits suggest that we imbue trust with intrinsic value too in the perception that broken trust doesn't just cost us utility but represents a 'wrong' committed against us *(*Ashraf, Bohnet and Piankov, 2006*).*

In summary, trust is beneficial to economic prosperity, individual well-being, social cohesion and political legitimacy.

2.4 Trust and Trustworthiness

The leading trust measures use surveys of public opinion - which are cost-effective, straightforward and intuitive (UN, 2021). These surveys are valuable as indicators of perception, but they only tell us how far people trust the institution, not whether that opinion is *justified*. This represents the right hand side of the arrow in Figure 1 but neglects the rest of the relation and its climate. Perception is subject to biases, heuristics and context, and when applied to perceptions of institutions, examples

[5] https://www.globalgoals.org/take-action/?id=1

include availability bias, representativeness and declinism (Frontiers[6]). Further, if perceptions are formed via online information, then salience, bandwagon effects, and homophily are documented issues in social media as is the prevalence of misinformation (Allcott, 2016; Lazer et al, 2018; Moravec et al, 2020; Pennycook et al, 2020,).

While behavioural interventions may seek to change perceptions, it cannot change the underlying institution and how worthy it is of public trust. If the intervention falsely inflates the perception of trust, beyond what is justified by the institution, dishonesty risks perpetrating a form of moral injury. If honesty is a key component of trust, then an institution using behavioural science to increase public perception of its trustworthiness beyond justified levels would be behaving dishonesty and therefore *less* worthy of trust. Should this come to public attention it would further damage trust perceptions. We contend that such an intervention would be unethical and fail the FORGOOD framework in respect of Fairness (by favouring the targeter over the target), Openness (lack of transparency) and Respect (undermining autonomy and dignity by seeking to mislead) (Lades and Delaney, 2022). Without an accurate measure of trustworthiness, it is difficult to objectively make such an assessment.

A developing theme in the literature has been the observation that while the perception of trust matters for the behaviour of the public, trustworthiness itself also matters if trust is to be lasting (Breuer and McDermott, 2010; Jones, 2012; O'Neill, 2018). There is an absence of attempts to measure trustworthiness in the literature, which may partly reflect the difficulties of measuring qualitative factors. Values are notoriously hard to measure - integrity and honesty are key components of trust but have no intrinsic numerical value and furthermore, some values and norms change over time or have cultural specificity. Other aspects of trust, such as reliability and competence are better suited to quantitative assessment such as measuring the extent to which stated goals and targets have been achieved.

One way to demonstrate trustworthiness is the relatively recent 'Market Failure Approach' (MFA) to business ethics. It builds upon the weaknesses of stakeholder theory and presents an alternative approach for interpreting what constitutes ethical behavioural science within the private sector (Heath, 2014). The MFA suggests that ethical corporations should behave as though market conditions are perfectly competitive, even though they may not be. It requires companies to avoid exploiting potential market failures in the way they conduct business. While such an approach to ethics can raise scepticism given the onus placed on corporations to act ethically by their own

[6] https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1157828/full

volition, adhering to the MFA can also be perceived to be consistent with the best long-term interests of shareholders. As the MFA is moral principle-led rather than explicitly prescriptive, it allows for adaptation and evolution. This is important and better suited to today's rapidly evolving business environment where legislation may not otherwise be able to keep pace.

3 Trust Surveys

Three surveys were studied as part of the present research: the Edelman barometer, the World Bank and the UN trust surveys. We have also looked at the recent report from the OECD, which assesses and compares perception of trust vs their behaviour in institutions. These surveys also contain useful additional demographic data.

Figures 2: YoY UN trust value

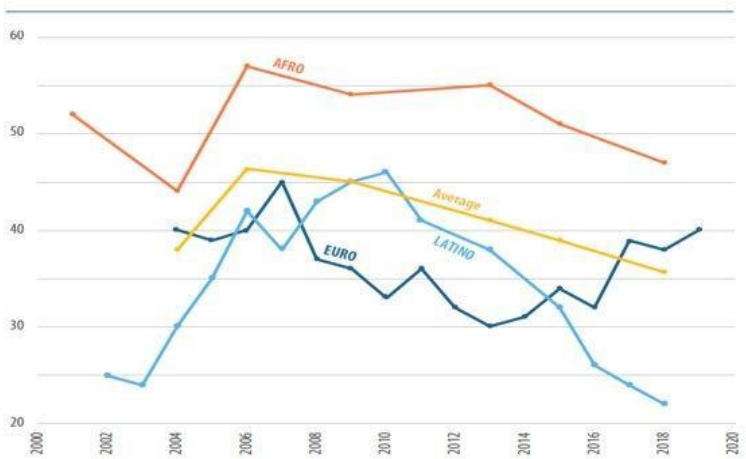
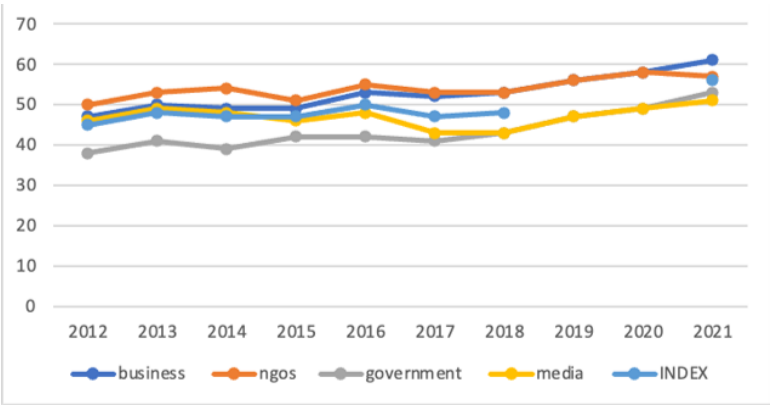


Figure 3: YoY Edelman trust barometer



These graphs are taken from the latest surveys conducted by the UN (2021) and Edelman (2022) respectively and although there are differences in survey design (the UN charts institutions broken down into four geographical regions, while Edelman averages trust scores across geographies but

charts results of four types of institution) there is a clear disparity between the trends with Edelman shows a rising trend whereas UN shows a declining trend. We have not seen a discussion within the literature that seeks to define or reconcile these conflicting results.

3.1 UN Survey Methodology

The UN combines data from the world values survey, the Gallop World poll and various regional barometers to compare surveyed trust measures across countries and identify trends in institutional trust. The methodology is described as "Author's calculations (unweighted averages) based on data from Eurobarometer, Afrobarometer and Latinobarometro. Eurobarometer figures are taken since 2004 EU expansion onwards to maximize countries included, while avoiding composition effects. They measure two values per year however only the earliest survey value in the year is selected. Afrobarometer data consists of only 6 waves, with waves covering 2/3 years of field work. The responses relate to "trust in parliament" (as there is no "trust in government" question on the survey. For Afrobarometer and Latinobarometer, graph combines "a lot" and "somewhat" into a single positive value and "none" and "not much" into a single negative response. Graph excludes "don't know/no response" and calculations only include country averages with continuous data, to avoid composition effects)"

A weakness of this approach is the absence of detail on the methodology making a direct comparison to other measures difficult and the absence of the rich supplemental data enabling a more nuanced understanding of dynamics at play. It also only measures trust in the national government. Comparable data going back to 2000 is a strength, as is the regional breakdown, but two major regions (US and Asia) are missing from the survey as their data does not go back as far.

The UN findings appear to show a broadly similar trend across different regions but with differences in the absolute levels of trust (much higher for Latino region, then the African region) and the individual data points within the overarching trends. Trust values for Africa and Latin America seem to spike upwards in 2003 and then trend downwards from 2006 onwards, while European trust levels spike upwards later, in 2006, then drift lower until 2014 when they appear to rise again.

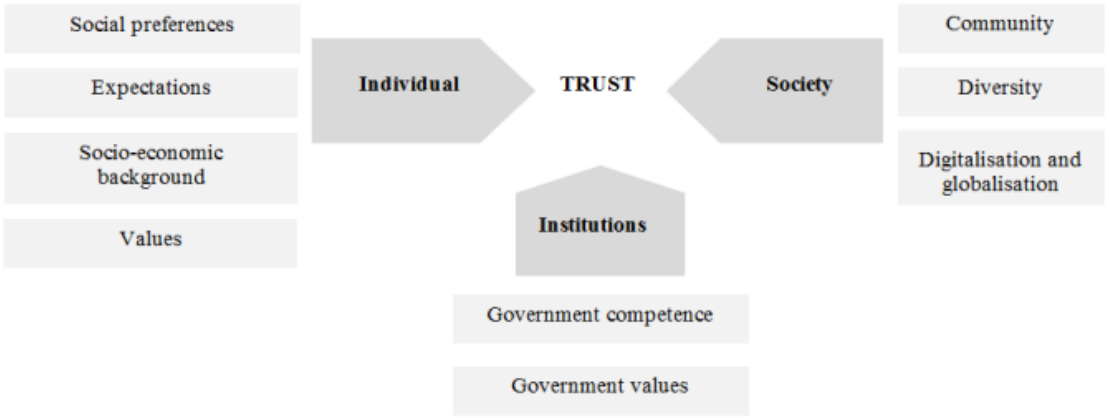
3.2 OECD Survey Methodology

The OECD uses a report from TrustLab to measure the perception of trust in government and fellow individuals using psychometric tests and self-reported survey data in 6 OECD countries globally (France, Germany, Italy, Korea, Slovenia and the United States). The data in each of the 6 OECD countries is collected using an online platform and completed by a minimum sample of 1,000 respondents. This sample is provided by a private-sector polling company and is nationally

representative by age, gender and income. Trustlab data collection method consists of three modules, which focus on behavioural games such as trust games, public good games, dictator and risk ladder games, eliciting measures of social norms, including trust in others and trustworthiness; an Implicit Association Test to capture implicit levels of trust in government and in the judicial system and, a traditional survey module with an extensive set of questions on interpersonal and institutional trust, as well as socioeconomic and demographic background variables. The survey also collected country-specific optional modules investigating specific topics of interest available eg. In the United States and Germany, a second trust game was included focused on 'bilateral' trust between different ethnic and racial groups (in the former) and between natives and migrants (in the latter); while in the case of Trustlab Italy questions on personality traits and fertility preferences were included.

The main policy determinants of (self-reported) trust in government are perceived government integrity, government reliability and government responsiveness, as well as satisfaction with certain public services, government fairness and perceptions of integration of immigrants. In the experiment, people's own behaviour predicts decisions about how much they can trust others, whereas the survey measure captures an inherent belief about others' trustworthiness. Figure 4 represents the OECD trust lab's conceptual framework for the determinants of trust. The self-reported survey data can then be compared to the experimental behaviour and the difference measured in behavioural and attitudinal perceptions of trust.

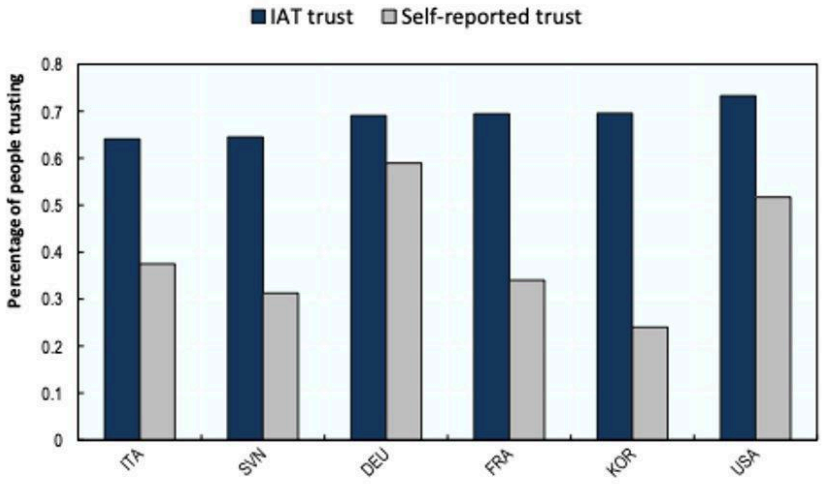
Figure 4: Conceptual framework for the determinants of trust according to OECD TrustLab



The TrustLab data suggest that there is a mismatch between an individual's behavioural and attitudinal perception of trust. The results suggest that people actually trust the institutions more

than they report they do. This brings us to think critically if behavioural perception is yet to catch the attitudinal perception or if social media has changed the attitudinal perception because of misinformation, homophily and echo chambers. Figure 5 represents the difference between self-reported data and experimental behavioural data.

Figure 5: OECD TrustLab comparison of behaviour and perception



The strength of Trustlab data is that it is the first internationally comparable and nationally representative data collection exercise on trust and other social preferences based on techniques from behavioural science and experimental economics, which allows comparison and a better understanding of both self-reported and experimental measures of trust. The experimental and survey measures represent different indicators of the same construct; Trustlab allows for the examination of validity. However, coverage is limited with data only collected for 6 countries which doesn't give a picture of global trends.

3.3 Edelman Survey Methodology

The Edelman Survey measures trust perception of government, media, NGOs and businesses in 28 countries, the results of which are equal-weighted to form an average trust measure. This annual online survey is conducted on a global basis for 3 weeks in November. Online panel participants are quota-matched for age, gender and region, and some countries are matched for ethnicity or nationality and asked to rate on a 9-point scale: "How far do you trust this institution to do what's right". Other characteristics are captured to enable additional analysis including household income, education level, employment status, the industry of work, job level and political party in the US. Expertise is also sought to translate questions as equivalently as possible.

The Edelman survey results highlight the gradual rise in the average score over the past 10 years obscures a growing divide between high and low-income participants - suggesting that trust is building for those who are prospering but declining for those who are not.

The strengths of Edelman's approach are that it uses a rich dataset that is genuinely global and grounded in the literature. It has been consistent over time, has considerable power (over 36,000 participants) and captures a great deal of information for supplementary analysis to compare countries, industries and comparison of different sections of the population. The four dimensions of trust used as a conceptual framework are fairness, vision, honesty and purpose-driven. Each institution is also charted according to two dimensions: competence and ethics. This is consistent with the conceptual framework outlined in section 2. It is also a valuable distinction to highlight the media and business as two major influences on our collective experience. Its weakness is that fully comparable data is only available since 2012 and that, as a commercial rather than academic undertaking, not all of the data is available for analysis. We understand that Edelman performs analysis of some environmental effects on perception but existing reports do not track these effects over time.

All three surveys appear methodologically sound and our analysis shows they were conducted with sufficient power. Every methodology has trade-offs, one common caveat being that self-reported online surveys may suffer from biases (such as selection bias, omitted variable bias and observation effects). It is also difficult to disaggregate the relational and environmental aspects or adjust to changing social expectations over time.

4 Key Findings and Discussion

We began with a literature review looking at what constitutes trust and how organisations measure trust. We then compared the differences in measurement techniques by Edelman, UN and OECD/TrustLab to identify the strengths and weaknesses of each technique, gaps in the literature and the reliability of each measurement approach.

As wide and deep as the term trust is, we realised there is no single parameter which can help us assess trust in institutions. In addition, the trends highlighted in Edelman and UN reports are not consistent. The average trend in Edelman shows that trust is increasing in institutions, while OECD and TrustLab show that trust is declining in institutions. After digging deeper into the data sets, we realised there is an ongoing trend in Edelman data which showcases that trust is increasing among highly educated and higher social strata individuals and is actually declining among poor and less educated individuals. This trend is alarming as it is leading to an increased gap between rich and poor and thus a constant threat to democracy.

Meanwhile the reliability of the survey measures is called into question by the TrustLab comparison between self-reported trust and behaviour. It remains an open question as to which is more reliable.

We also tried to understand the psychological drivers for trust and what drives people's trust in institutions. The literature from social identity suggests that trust is a factor of in-group bias and homophilly (OECD)[7] An individual's propensity to trust increases if the government/ institutions make policies in favour of their citizens and for their wellbeing. This aligns with the components of trust identified earlier, including competence, ethical values, integrity, and motivations .

5 Further Research

We have not examined the cultural dimension to trust in this paper and this would be an area for future research. The Edelman results can be viewed on a country-by-country basis and show some sharp divergence, but deeper research would be required to identify specific cultural drivers. Section 2.2 argues that the climate, or environment in which the trust relation exists has an impact on the quality and degree of trust. If that understanding is correct, country-specific factors may play a role by forming part of that environment.

Another environmental aspect we have not explicitly addressed here is the ubiquity of technology, presenting new ways of interacting with institutions, consuming information and of forming opinions. A recent survey suggests that trust is highest where it is closest to us – our immediate circle of friends and family or our employer. (Edelman, 2022). New structures may be needed to enable citizens to assess the trustworthiness of their institutions (O'Neill, 2018). However, the absence of data prior to 2000, makes comparisons with a pre-technological, internet-enabled world difficult. Technology also offers new techniques for identifying public mood and implied beliefs, such as textual analysis, although to our knowledge this is yet to be deployed in a meaningful way when measuring trust perception.

[7] https://doi.org/10.1787/869ef2ec-en

6 Conclusion

We propose that an objective measure of trustworthiness is an essential prerequisite to being able to develop policy relating to trust in institutions and to identify appropriate, effective, and ethical interventions. If an institution designed an intervention to improve trust perception when the perception is actually accurate, that would be dishonest and risk moral injury. The harm emanating from any resulting backlash would not be limited to further declines in trust in that institution, it could impact perceptions of behavioural science more generally. Developing a trustworthiness measure would also be beneficial in improving the accuracy of trust perceptions and provide wider context and benchmarking when assessing individual institutions. Having context can better enable transparency to have a confidence-building effect rather than just increasing the salience of bad news. Furthermore, new tools have the potential to address this shortfall.

Overall, trust perceptions do seem to be lower - but the average scores shown in 3.1 mask a more complicated underlying picture, in which there is considerable variation across geography and socioeconomic status. Emerging analysis from the OECD suggests that behaviour and perception may be misaligned, which casts some doubt on taking the output of surveys at face value, and it would be useful to better understand this relationship in more detail and whether environmental factors play a role. There are many behavioural aspects that could cause inaccurate perceptions of trust, including salience, Availability Heuristic, Representative Heuristic, and Present bias. There has also been a profound change in environmental factors such as the way in which we interact with institutions and how we consume news and information. This indicates a wealth of prospects for developing behavioural interventions once other concerns have been resolved.

7 References

Allcott. H., and Gentzkow. M, (2016) *Social Media and Fake News in the 2016,* Election Journal of Economic Perspectives, Volume 31, Number 2, Spring 2017, Pages 211–236

Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental economics*, *9*(3), 193-208.

Barney, J. B., & Hansen, M. H. (1994). Trustworthiness as a source of competitive advantage. *Strategic management journal*, *15*(S1), 175-190

Bohnet, Iris, and Steffen Huck. 2004. "Repetition and Reputation: Implications for Trust and Trustworthiness When Institutions Change." *American Economic Review*, 94 (2): 362-366.

Bowles, S., & Polanía-Reyes, S. (2013). Economic Incentives and Social Preferences: Substitutes or Complements? (Part 1). Voprosy Ekonomiki, 4, 24–48. https://doi.org/10.32609/0042-8736-2013-4-24-48

Breuer and McDermott, *Trustworthiness and Economic Performance* (August 17, 2010). Available at SSRN: https://ssrn.com/abstract=1314844

Brennan, 2021 Recognition Trust, Philosophical Studies, (2021) 178:3799–3818

Chang, Dillon, Hussain, 2006, *Trust and Reputation for Service-Oriented Environments:* Technologies for Building Business Intelligence and Consumer Confidence

Kidwell, M. C., & Kerig, P. K. (2021). To trust is to survive: Toward a developmental model of moral injury. *Journal of Child & Adolescent Trauma*, 1-17

Lades, L., & Delaney, L. (2022) Nudge FORGOOD. *Behavioural Public Policy, 6*(1), 75-94. Available at: doi:10.1017/bpp.2019.53

Lazer. D. M. J., Baum. M. A., Benkler. Y., Berinsky. A. J, Greenhill. K. M., Menczer. F. Metzger. M. J. Nyhan. B. Pennycook. G. Rothschild. D. Schudson. M. Sloman. S. A, Sunstein. C. R. Thorson**.** E. A, Watts. D. J & Zittrain. J. L (2018) *The Science of Fake News,* SCIENCE, 9 Mar 2018, Vol 359, Issue 6380, pp. 1094-1096. Available at: DOI: 10.1126/science.aao2998

McLeod, Carolyn, "Trust", *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), <https://plato.stanford.edu/archives/fall2023/entries/trust/>.

Moravec, P.L., Kim. A., and Dennis. D. R, (2020) Appealing to Sense and Sensibility: System 1 and System 2 Interventions for Fake News on Social Media Information Systems Research 2020 31:3, 987-1006

O'Neill, 2018, *Linking Trust to Trustworthiness* International Journal of philosophical studies**,** Volume 26, 2018, Issue 2 **https://doi.org/10.1080/09672559.2018.1454637**

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2020). *Understanding and reducing the spread of misinformation online.* PsyArXiv Preprints. doi:10.31234/osf.io/3n9u8

Shay, J. (2014). Moral Injury. *Psychoanalytic Psychology, 31* (2), 182-191. doi: 10.1037/a0036090.

Song & Lee (2016) *Citizens' Use of Social Media in Government, Perceived Transparency, and Trust in Government*, Public Performance & Management Review, 39:2, 430-453

Sucher and Gupter, 2022, *The power of trust*, public affairs books, 2022

Wang, 2015, *Government Performance, Corruption, and Political Trust in East Asia*, Social Science Quarterly, Volume 97, Issue 2, June 2016, Pages 211-31

"Trust and Its Determinants." OECD Statistics Working Papers, 2018, https://doi.org/10.1787/869ef2ec-en.

7.1 Other Resources:

Online resource for the Moral Injury Project at Syracuse University
https://moralinjuryproject.syr.edu/about-moral-injury/

UN Sustainable Development Goals website: https://sdgs.un.org/goals/goal16

EU sponsored report on Trust
https://www.tigre-project.eu/members/repository/Public/Deliverables/TiGRE_D1.1_v1.0.pdf

OECD report on trust in government: https://data.oecd.org/gga/trust-in-government.htm

UN report on trust in institutions:
https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2021/08/PB_108.pdf

Edelman annual trust report: https://www.edelman.com/trust/2022-trust-barometer

World Bank working paper on Building Trust in Government Through Citizen Engagement:
http://hdl.handle.net/10986/33346

Murtin. F, Fleischer. L, and Siegerink. V (2018) 'Trust and its determinants: Evidence from the Trust Lab Experiments'. *OECD working paper: OECD Statistics and Data Directorate,* Available at:

https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=SDD/DOC(2018)2&docLanguage=En

Applied behavioural insights consultancy and resource centre: https://thedecisionlab.com/biases

Behavioural Science Guide 2022:
https://www.behavioraleconomics.com/be-guide/the-behavioral-economics-guide-2022/

World Values Survey (KCL, Behavioural Insights Team, The Policy Institute et al) 2023

https://www.kcl.ac.uk/policy-institute/assets/confidence-in-institutions.pdf

7.2 Acknowledgements:
The authors would particularly like to thank the Edelman team for their time and engagement.

8 APPENDICES

8.1 Potential Intervention

If it becomes possible to develop a reliable measure of trustworthiness that would justify an ethical intervention, to correct a discrepancy between actual and perception, some prospects for such an intervention are discussed below. If the government is trustworthy but the perception of trust is low among individuals, then an appropriate behavioural intervention can be deployed to increase trust. One way to increase process-based trust, which will eventually lead to increase in overall political trust, will be to increase the level, depth and quality of engagement with citizens by institutions.

There have been huge changes to the social, media and information environment, since the development of the internet, which may partly explain changing perceptions. How we consume news and information has changed profoundly in the last 20 years and has therefore changed the environment in which the relation of trust exists between person and institution. (Sunstein, 2017; O'Neill, 2018) While it may not be possible to uninvent the internet, a source of future research could look at how far institutions have changed their behaviour to reflect that changed environment and whether an intervention could have an impact.

One of the aspects we would like to investigate further would be identifying how best to enable individuals to form an accurate perception of how justified any institution is when seeking to be trusted. If an institution acknowledges wrongdoing in a public statement but is otherwise relatively quiet in public perception, an inaccurate perception could form that this is representative of the institution rather than an aberration. The availability heuristic, representative heuristic, salience and framing could all contribute as behavioural aspects that could result in a mistaken impression forming. Developing objective measures to help provide a framing reference could be useful in countering those heuristics.

8.2 Trustworthiness Measures for further development

A measure to rate the trustworthiness of institutions over time would face the typical difficulties of seeking to quantitatively score complex qualitative issues. Experimenting with different combinations of features would need to be an iterative process. We propose testing the features outlined below and incorporating a qualitative overlay to allow for extenuating factors.

8.2.1 Country Context

The country context is part of the environment in which trust is formed and cultural norms are situated. There are numerous existing measures which provide insights as to strengths and weaknesses of institutions within particular countries including: the press transparency index, corruption index, degree to which elections are free and fair, and risk of money laundering.[8] Assessing the rigour and independence of the legal system, as well as the strength and independence of regulatory bodies could also be instructive.

8.2.2 Institutional Assessment

There are several indicators which can be used to assess the governance of individual institutions that are outlined in guidance papers from independent bodies such as the Institute of Business Ethics or Financial Conduct Authority. Within the commercial sector these are often considered as part of Operational Due Diligence and by external or internal audit. We would propose an iterative testing process that includes considerations such as the quality of an institution's code of ethics and whistleblower provision, reviewing instances of regulatory fines or legal censure and official complaints as well as typical organisational culture indicators including staff turnover, use of NDAs and analysis of social media.

8.2.3 Use of Technology

Finally, we note that Artificial Intelligence has facilitated the gathering and processing of information in new ways and there are numerous examples of measures of textual analysis using Natural Language Processing (NLP) that can scan public statements and social media representation to identify staff, supplier and customer observations, any trends towards deterioration or improvement, staff and an assessment of how far stated promises have been kept and goals achieved.

8.3 Assessment of Environmental factors:

An assessment of the environmental context in which individuals perceive the trustworthiness of institutions, and in which institutions seek to build trust amongst the public will be informative to better understand the drivers of trust and potential for interventions. To understand environmental effects, we would like to do a detailed correlation analysis and invite reviewers to critique and expand our analysis. We propose including factors such as the unemployment rate, real terms wage growth and GDP growth as part of an economic environmental assessment. Political environment factors could include the stability of government by stage in the electoral cycle and strength of any majority, as well as external review of the fairness of elections. Access to information is another environmental effect we would like to research further, including such factors as the freedom of the

[8] https://www.transparency.org/en/press https://www.transparency.org/en/cpi/2021 https://ourworldindata.org/grapher/free-and-fair-elections?country=ARG~AUS~BWA~CHN https://www.fatf-gafi.org/

press, access to the internet, degree of censorship, measures of degree of misinformation or disinformation. Finally we would like to assess the effect of happiness on levels of trust by regressing country data against the Global happiness index. [9]

[9] https://worldhappiness.report/news/media-round-up-on-the-world-happiness-report-2022/