What Phishing Click Rates Don't Show

Sarah Y. Zheng[1,2] and Ingolf Becker[2]

[1] Dawes Centre for Future Crime
[2] Department of Security & Crime Science, UCL
{sarah.zheng.16,i.becker}@ucl.ac.uk

Abstract. High-profile ransomware incidents, such as the April 2025 attack on Marks & Spencer, highlight a persistent paradox in cybersecurity: major organisations still suffer severe disruption following a single human error. This paper argues that the prevailing focus on phishing click rates as a primary indicator of human cybersecurity risk is both conceptually flawed and unnecessarily costly. Drawing on psychological research, we show that humans are inherently poor at deception detections and that a non-zero baseline of phishing victimisation is unavoidable, regardless of technological literacy or how often employees complete mandatory cybersecurity trainings. Framing cybersecurity performance around click rates encourages blame, erodes trust, and obscures deeper organisational and systemic risk factors. We propose a shift from prevention-centric metrics toward holistic human risk management that emphasises behavioural indicators, organisational culture, psychological safety, and response capacity. By embedding cybersecurity within business continuity planning and focusing on resilience, organisations can better anticipate, manage, and recover from inevitable human security failures.

Keywords: phishing · human risk · cybersecurity resilience · organisational security · social engineering

1 Introduction

In April 2025, major British retailer Marks & Spencer (M&S) was hit by a ransomware attack [19]. The intrusion began with a phishing e-mail that targeted a third-party vendor's access credentials, which allowed the attackers to infiltrate M&S' systems. The incident caused significant operational disruption and financial losses, with online services offline for weeks, substantial impacts on sales, and hundreds of millions lost in market value [25]. Events like these raise the question how an organisation like M&S could suffer such lasting disruptions after one individual entered their credentials on a phishing website. It would certainly be easy to blame this person and fire them, but that would negate more deeply

rooted issues with how organisations secure their information technology (IT) landscape and manage human risk.

Especially the latter is reflected in an obsession over reducing phishing click rates—both in industry, and academic research on interventions seeking to improve cybersecurity behaviour. In most cases, these entail some form of internal phishing simulation or cybersecurity education program. Ambitious organisations may implement further cybersecurity awareness-raising activities throughout the year or even award employees that become cybersecurity champions [26]. Yet, ultimately, executive boards will only care about green compliance checks and getting near-zero phishing click rates [29]. Phishing click rates have become the default measure of human cybersecurity risk, precisely because they are easy to quantify. This is an understandable, but costly mistake. Indeed, according to Goodhart's Law, "when a measure becomes a target, it ceases to be a good measure" [11].

Cybersecurity training and education companies capitalise on this single metric. They will illustrate the constantly looming threat of phishing attacks and show dramatic drops in phishing click rates after organisations implement their training program. However, a growing number of independent studies shows that training effects from both simulated phishing tests and cybersecurity awareness programs are modest, inconsistent, and mostly fail to persist [31,23,22]. A small, but persistent percentage of people will click on (simulated) phishing links, no matter the training or phishing simulation frequency [9]. Organisations would be right to question whether these mandatory programmes are worth the costs: not only financially, but also in terms of the annoyance and distrust they can cause among employees [6,36]. Compromise thus becomes not a possibility, but an expected reality in due course—because phishing and other social engineering attacks are not new. They are scams as old as humanity. Networking and AI technologies have only enabled fraudsters to target a much larger number of people than before. Modern anti-phishing technology has reduced this number, but it has only forced fraudsters to return to more sophisticated and targeted impersonation attacks. This demands a shift in perspective from an exclusive focus on prevention from within the intuition behind monitoring phishing simulation click rates, toward resilience and recovery. Organisations must assume that breaches will occur due to the inevitable possibility that someone in their organisation, whatever their tenure, can at some point fall for a phishing e-mail.

Hence, this paper argues that phishing click rates are not merely imperfect proxies for human cyber risk, but that their use as a primary performance indicator actively undermines effective risk management. Click rates capture isolated moments of failure without accounting for the psychological, situational, and organisational conditions under which people work. When formalised as key performance indicators (KPIs), they encourage individual attribution of blame, erode psychological safety, and obscure systemic contributors to risk, such as workload pressures, ambiguous authority structures, and poorly designed technical systems. As a result, organisations optimise for the appearance of prevention—and not for resilience.

Instead of pursuing the elimination of clicks (i.e., an unattainable goal), organisations must assume a non-zero baseline of human error and design cybersecurity measures accordingly. This requires a shift from prevention-centric metrics toward a holistic approach to human cyber risk management that emphasises behavioural patterns, organisational culture, improving its response capacity, and recovery. Cybersecurity must therefore be embedded within business continuity planning and organisational strategy, recognising that failure is inevitable, but large-scale disruption is not [34].

2 The real human risk factors

A substantial body of psychological research demonstrates that humans are generally poor at detecting deception [13,8,15]. People tend to rely on contextual cues and heuristics that are easily manipulated, while underutilising more reliable indicators of dishonesty. This limitation is not a function of intelligence, training, or technical expertise: individuals across age groups, professions, and seniority levels are vulnerable to well-crafted deception (e.g., [39,32,38]). It is therefore unsurprising that phishing attacks continue to succeed, even in organisations with extensive awareness programmes.

In everyday work contexts, people prioritise efficiently completing their primary tasks [5]. When managing email, they typically attend to cues that support this goal, such as perceived legitimacy, urgency, and relevance, rather than scrutinising technical security indicators [40]. Only when something appears sufficiently unexpected or anomalous do individuals shift their attention toward security checks such as inspecting sender addresses or hovering over links [38,40]. Even then, many users lack the skills or confidence to accurately interpret these signals [2,37]. Phishing detection is therefore intermittent, effortful, and highly sensitive to context.

Situational factors can further exacerbate vulnerability. Time pressure, cognitive overload, stress, and fatigue all degrade attention and increase reliance on heuristics. Under such conditions, even highly experienced and security-conscious individuals may act impulsively. A non-zero base rate of phishing victimisation is therefore not a training failure, but an expected outcome of human cognition operating under real-world constraints.

Attempts to identify and manage "repeat clickers" illustrate the limitations of a click-rate-focused approach. While a small percentage of individuals may repeatedly fall for simulated phishing emails [9,10], the underlying drivers of this behaviour are poorly understood and may relate to psychological traits, such as how honest people believe others are in general [10,42] and situational job demands. Treating repeated clicks as evidence of negligence or incompetence risks misdiagnosing structural problems as individual deficiencies.

In this paper, human cyber risk is understood not as an individual propensity for error, but as the probability that organisational conditions will place people in situations where error is both likely and consequential. From this perspective, human risk emerges from the interaction between cognitive limitations, situa-

tional pressures, and system design. In fact, one could argue that cybersecurity training for the general user is necessary, precisely because IT systems are suboptimal [1]. Users are forced to apply password complexity rules to compensate for weak hashing algorithms in legacy Windows systems (e.g., the lack of salting in NTLM, flaws in LM Hash [27]). Similarly, users need to update their passwords before an expiry date, because systems cannot detect when a password actually is compromised, despite sources like https://haveibeenpwned.com/. Many organisations have also struggled setting up simple SPAM filters [24], requiring individuals to stay vigilant when using computers. Managing human cyber risk effectively therefore requires interventions that go beyond awareness training and address how work actually is organised [34,1,16].

A further consequence of reducing human cyber risk to click rates is the emergence of blame-oriented cultures. Once click rates are formalised as performance indicators, they become tools for comparison, sanction, and reputational judgement. Managers may seek to discipline "poor performers", human resources may associate security failures with broader performance issues, and security teams may prioritise reducing reported clicks over improving reporting behaviour. This dynamic discourages openness, delays incident escalation, and ultimately increases organisational risk.

By contrast, research on psychological safety shows that environments in which individuals feel able to report mistakes without fear of punishment enable faster detection and containment of incidents. Blameless postmortem practices from DevOps and Site Reliability Engineering offer valuable lessons [17]: learning-oriented responses to failure improve system reliability precisely because they focus on conditions, instead of culprits.

3 From human risk to cyber resilience

If phishing click rates obscure rather than illuminate human cyber risk, organisations require alternative ways of seeing, interpreting, and managing human behaviour under realistic conditions. They need to ask questions like what is our human risk? What does manageable human risk look like in our specific industry, sector, and region? What digital norms do we want to set and expect for our employees? How do we want to deal with unsolicited requests? How do our current communication processes and policies allow for human risk? How can we tailor cybersecurity training and awareness programs to people's specific job contexts?

This is not to say that compliance is redundant. Regulatory requirements, liability protection, insurance prerequisites, and contractual obligations all drive compliance behaviour. But compliance typically takes root in a prevention mindset which can provide a secure illusion. Organisations optimise for cost, and prevention *feels* cheaper than resilience investment—until it is not. The good news is that measures focused on people tend to be cheaper than technical ones. Below we outline a set of eight human-centric recommendations that shift the focus from individual prevention toward organisational resilience, where resilience

refers to the capacity to anticipate, absorb, recover from, and adapt to cyber incidents—not the elimination of failure.

Identify behavioural indicators beyond phishing simulations. Simulated phishing click rates offer, at best, a narrow snapshot of human cyber risk. They capture isolated moments of failure while ignoring broader behavioural patterns that shape everyday security posture. A more informative approach involves monitoring indicators such as excessive or misaligned file access [7], habitual rapid link-clicking, repeated bypassing of security prompts (although see [3] for a recent review with recommendations on how to implement better user warnings), unsanctioned external data exports under social pressure [28], or risky policy-violating searches using public search engines or generative AI tools.

These behaviours often reflect time pressure, cognitive overload, or misaligned workflows rather than malicious intent. When interpreted proportionately and ethically, they can help organisations identify systemic friction points where employees are forced to trade security for productivity. Unlike click rates, such indicators could support targeted improvements in system design and organisational policy rather than punitive retraining.

Advance from awareness training to adversarial training. Traditional cybersecurity training assumes that awareness leads to secure behaviour and that individuals make decisions under calm, rational conditions. However, real-world attacks violate both these assumptions. Social engineering attacks exploit stress, authority gradients, emotional manipulation, and fatigue. Victims do not necessarily act because they underestimate risk, but because they overestimate the consequences of non-compliance with the contents of a social engineering attack.

Adversarial training reframes cybersecurity education by focusing on how attackers think, adapt, and exploit organisational dynamics [41]. Rather than teaching employees how to spot suspicious emails, it exposes them to realistic scenarios involving urgency, ambiguity, and conflicting priorities. The goal is not perfect prevention, but earlier recognition, escalation, and recovery when something feels odd. Although this training paradigm requires further empirical support, the expectation is that aligning training with employees' actual job contexts will be more effective for protecting them from the psychological dimensions of cyber threats.

Establish shared digital norms for ambiguous and urgent requests. Many successful cyber incidents exploit uncertainty rather than technical vulnerability. Employees are confronted with requests that appear urgent, authoritative, or only slightly irregular [12], and must decide whether caution will be rewarded or penalised. Organisations can reduce this ambiguity by establishing explicit digital norms for handling unsolicited or high-pressure requests. Such norms may include mandatory secondary verification channels, pre-agreed response scripts, or explicit permission to delay action. Crucially, leadership must model these behaviours. When senior figures accept verification without frustration, security becomes a collective expectation rather than an individual risk.

Build psychological safety and manage stress as security controls. Organisational culture is a critical but underutilised security control. Blame-oriented responses discourage reporting and delay containment [14,30]. Psychologically safe environments, by contrast, enable early disclosure and rapid response. Stress and burnout further amplify risk by increasing impulsivity and encouraging insecure workarounds. Treating wellbeing and workload management as integral components of cybersecurity recognises such cognitive limits.

Integrate and test cyber resilience as part of business continuity planning. Whereas traditional business continuity management assumes that systems can be meaningfully decomposed and that risks can be predicted, this defensive posture provides an illusion of control that collapses under real-world uncertainty [35]. A resilience engineering approach instead designs for recovery from the outset. Just as infrastructure-as-code tools like Terraform enable reproducible system provisioning, and data protection regulations such as the GDPR's right to data portability mandate machine-readable exports, organisational processes should be architected bottom-up with recovery as a first-class requirement. This means not merely having backups, but ensuring that restoration is tested, practised, and achievable within acceptable time frames [4]. Otherwise, incident response and disaster recovery plans are only as good as the last time they were exercised (spoiler: never).

Understand failing IT systems as organisational reflections. Technical systems do not fail in isolation. Under-resourcing, compressed timelines, and conflicting incentives often produce systems that allow for significant human risk by design. When security controls obstruct work, non-compliance becomes the rational choice [21,20]. Improving human cyber risk management therefore requires investment not only in technology, but also in the organisational conditions under which systems are built and maintained.

Reframe phishing metrics within a holistic resilience strategy. Phishing simulations need not be abandoned, but their role must be reframed. Click rates are most useful as diagnostic signals when interpreted alongside other behavioural indicators and organisational context. When treated as performance targets, they distort incentives; when treated as learning tools, they can inform system-level improvements.

Position insurance and recovery funds as a final defence layer. Cyber insurance and recovery funds should support recovery after failure, and not substitute robust security practices. Before relying on insurance, organisations must ensure foundational controls are in place [34]. In many cases, investing in usable security measures, such as passwordless authentication or physical security keys, reduces human error more effectively than complex policy enforcement [33,18].

4 Conclusion

Holistic cyber resilience strategies that emphasise anticipation, detection, recovery, and adaptation offer a more realistic path than compliance regimes built around prevention metrics. As long as organisations treat phishing clicks as failures to be eliminated rather than signals to be interpreted, they will continue to invest in the appearance of security rather than its substance. Sustainable cybersecurity emerges not from perfect defences, but from resilient systems and cultures that acknowledge human behaviour as central to cyber risk and are prepared to respond constructively when failure occurs.

References

1. Adams, A., Sasse, M.A.: Users are not the enemy. Communications of the ACM **42**(12), 40–46 (1999)
2. Albakry, S., Vaniea, K., Wolters, M.K.: What is this url's destination? empirical evaluation of users' url reading. In: Conference on Human Factors in Computing Systems - Proceedings. Association for Computing Machinery (4 2020). https://doi.org/10.1145/3313831.3376168
3. Amran, A., Zaaba, Z.F., Singh, M.K.M.: Habituation effects in computer security warning. Information Security Journal: A Global Perspective **27**(2), 119–131 (2018). https://doi.org/10.1080/19393555.2018.1448492, https://doi.org/10.1080/19393555.2018.1448492
4. Angafor, G.N., Yevseyeva, I., Maglaras, L.: Scenario-based incident response training: lessons learnt from conducting an experiential learning virtual incident response tabletop exercise. Information & Computer Security **31**(4), 404–426 (2023). https://doi.org/10.1108/ICS-05-2022-0085
5. Bada, M., Sasse, A., Nurse, J.: Cyber security awareness campaigns: Why they fail to change behavior. International Conference on Cyber Security for Sustainable Society (7), 38 (2014), http://www.cs.ox.ac.uk/publications/publication9343-abstract.html
6. Beautement, A., Sasse, M.A., Wonham, M.: The compliance budget: managing security behaviour in organisations. In: Proceedings of the 2008 new security paradigms workshop. pp. 47–58 (2008). https://doi.org/10.1145/1595676.1595684
7. Bertino, E., Ghinita, G.: Towards mechanisms for detection and prevention of data exfiltration by insiders: keynote talk paper. In: Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security. p. 10–19. ASIACCS '11, Association for Computing Machinery, New York, NY, USA (2011). https://doi.org/10.1145/1966913.1966916, https://doi.org/10.1145/1966913.1966916
8. Bond, C., DePaulo, B.M.: Accuracy of deception judgments. Personality and Social Psychology Review **10**(3), 214–234 (2006). https://doi.org/10.1207/s15327957pspr1003_2
9. Canham, M.: Repeat clicking: A lack of awareness is not the problem. In: Degen, H., Ntoa, S., Moallem, A. (eds.) HCI International 2023 – Late Breaking Papers. pp. 325–342. Springer Nature Switzerland, Cham (2023)

10. Canham, M., Posey, C., Strickland, D., Constantino, M.: Phishing for long tails: Examining organizational repeat clickers and protective stewards. Sage Open **11**(1), 2158244021990656 (2021). https://doi.org/10.1177/2158244021990656
11. Chrystal, K.A., Mizen, P.D., Mizen, P.: Goodhart's law: its origins, meaning and implications for monetary policy. Central banking, monetary theory and practice: Essays in honour of Charles Goodhart **1**, 221–243 (2003)
12. De Bona, M., Paci, F.: A real world study on employees' susceptibility to phishing attacks. In: Proceedings of the 15th International Conference on Availability, Reliability and Security. pp. 1–10 (2020)
13. DePaulo, B., Lindsay, J.J., Malone, B., Muhlenbruck, L., Charlton, K., Cooper, H.: Cues to deception. Psychological bulletin **129**, 74–118 (2 2003). https://doi.org/10.1037/0033-2909.129.1.74
14. Ebert, N., Schaltegger, T., Ambuehl, B., Geppert, T., Trammell, A., Knieps, M., Zimmermann, V.: Learning from safety science: designing incident reporting systems in cybersecurity. Journal of Cybersecurity **11**(1), tyaf019 (08 2025). https://doi.org/10.1093/cybsec/tyaf019
15. Hartwig, M., Bond, C.: Lie detection from multiple cues: A meta-analysis. Applied Cognitive Psychology **28**(5), 661–676 (2014). https://doi.org/10.1002/acp.3052
16. Hielscher, J., Schöps, M., Menges, U., Gutfleisch, M., Helbling, M., Sasse, M.A.: Lacking the tools and support to fix friction: results from an interview study with security managers. In: Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023). pp. 131–150 (2023)
17. Hoeppner, F., Sbaraglia, F.: Culture Shift for SRE Adoption, pp. 131–143. Apress, Berkeley, CA (2025). https://doi.org/10.1007/979-8-8688-1448-8_4
18. Inglesant, P.G., Sasse, M.A.: The true cost of unusable password policies: password use in the wild. In: Proceedings of the sigchi conference on human factors in computing systems. pp. 383–392 (2010)
19. Johnson, G.: Inside the m&s cyberattack: Timeline, https://www.the-web-people.com/blog/inside-the-ms-cyberattack-technical-analysis, accessed on 28 December 2025
20. Kirlappos, I., Parkin, S., Sasse, A.: Learning from "shadow security:" why understanding non-compliant behaviors provides the basis for effective security (02 2014). https://doi.org/10.14722/usec.2014.23007
21. Kirlappos, I., Sasse, M.A.: What usable security really means: Trusting and engaging users. In: Tryfonas, T., Askoxylakis, I. (eds.) Human Aspects of Information Security, Privacy, and Trust. pp. 69–78. Springer International Publishing, Cham (2014)
22. Lain, D., Jost, T., Matetic, S., Kostiainen, K., Čapkun, S.: Content, nudges and incentives: A study on the effectiveness and perception of embedded phishing training. In: Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security. pp. 4182–4196 (2024). https://doi.org/10.1145/3658644.3690348
23. Lain, D., Kostiainen, K., Čapkun, S.: Phishing in organizations: Findings from a large-scale and long-term study. In: 2022 IEEE Symposium on Security and Privacy (SP). pp. 842–859. IEEE (2022). https://doi.org/10.1109/SP46214.2022.9833766
24. Jáñez Martino, F., Alaiz-Rodríguez, R., González-Castro, V., Fidalgo, E., Alegre, E.: A review of spam email detection: analysis of spammer strategies and the dataset shift problem. Artif. Intell. Rev. **56**(2), 1145–1173 (May 2022). https://doi.org/10.1007/s10462-022-10195-4

25. Masud, F.: M&S profits almost wiped out after cyber hack hit sales, https://www.bbc.com/news/articles/c93x16zkl9do, accessed on 28 December 2025
26. Menges, U., Hielscher, J., Kocksch, L., Kluge, A., Sasse, M.A.: Caring not scaring-an evaluation of a workshop to train apprentices as security champions. In: Proceedings of the 2023 European Symposium on Usable Security. pp. 237–252 (2023)
27. Microsoft: Passwords technical overview. https://learn.microsoft.com/en-us/windows-server/security/kerberos/passwords-technical-overview (2024), accessed: 2026-01-07
28. Moore, L., Mori, T., Hasegawa, A.A.: Negative effects of social triggers on user security and privacy behaviors. In: Twentieth Symposium on Usable Privacy and Security (SOUPS 2024). pp. 605–622 (2024)
29. Opdenbusch, J., Hielscher, J., Sasse, M.A.: "Where Are We On Cyber?" A Qualitative Study On Boards' Cybersecurity Risk Decision Making. In: NDSS (2025). https://doi.org/10.14722/ndss.2025.240595
30. Patterson, C.M., Nurse, J.R., Franqueira, V.N.: "I don't think we're there yet": The practices and challenges of organisational learning from cyber security incidents. Computers & Security **139**, 103699 (2024). https://doi.org/10.1016/j.cose.2023.103699
31. Reinheimer, B., Aldag, L., Mayer, P., Mossano, M., Duezguen, R., Lofthouse, B., von Landesberger, T., Volkamer, M.: An investigation of phishing awareness and education over time: When and how to best remind users. Proceedings of the 16th Symposium on Usable Privacy and Security (SOUPS 2020) pp. 259–284 (2020), https://www.usenix.org/conference/soups2020/presentation/reinheimer
32. Sarno, D.M., Lewis, J.E., Bohil, C.J., Neider, M.B.: Which phish is on the hook? phishing vulnerability for older versus younger adults. Human Factors **62**(5), 704–717 (8 2020). https://doi.org/10.1177/0018720819855570
33. Sasse, M.A., Steves, M., Krol, K., Chisnell, D.: The great authentication fatigue – and how to overcome it. In: Rau, P.L.P. (ed.) Cross-Cultural Design. pp. 228–239. Springer International Publishing, Cham (2014)
34. Singh, T., Zheng, S.Y.: The Psychology of Cybersecurity: Hacking and the Human Mind. Taylor & Francis (2025)
35. Steen, R., Haug, O.J., Patriarca, R.: Business continuity and resilience management: A conceptual framework. Journal of Contingencies and Crisis Management **32**(1), e12501 (2024). https://doi.org/10.1111/1468-5973.12501
36. Volkamer, M., Sasse, M.A., Boehm, F.: Analysing simulated phishing campaigns for staff. In: Computer Security: ESORICS 2020 International Workshops, DETIPS, DeSECSys, MPS, and SPOSE, Guildford, UK, September 17–18, 2020, Revised Selected Papers. p. 312–328. Springer-Verlag, Berlin, Heidelberg (2020). https://doi.org/10.1007/978-3-030-66504-3_19
37. Wash, R.: Folk models of home computer security. ACM International Conference Proceeding Series (2010). https://doi.org/10.1145/1837110.1837125
38. Wash, R.: How experts detect phishing scam emails. Proceedings of the ACM on Human-Computer Interaction **4**(2) (10 2020). https://doi.org/10.1145/3415231
39. Zheng, S.Y., Becker, I.: Presenting suspicious details in User-Facing e-mail headers does not improve phishing detection. In: Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022). pp. 253–271. USENIX Association, Boston, MA (8 2022), https://www.usenix.org/conference/soups2022/presentation/zheng
40. Zheng, S.Y., Becker, I.: Checking, nudging or scoring? evaluating e-mail user security tools. In: Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023). pp. 57–76. USENIX Association, Anaheim, CA (Aug 2023), https://www.usenix.org/conference/soups2023/presentation/zheng

41. Zheng, S.Y., Becker, I.: Phishing to improve detection. In: The European Symposium on Usable Security (2023). https://doi.org/10.1145/3617072.3617121
42. Zheng, S.Y., Rozenkrantz, L., Sharot, T.: Poor lie detection related to an underreliance on statistical cues and overreliance on own behaviour. Communications Psychology **2**(1), 21 (2024). https://doi.org/10.1038/s44271-024-00068-7