

Enhancing corporate accountability through covert situational integrity testing (CSIT)

Nicholas Lord & Michael Levi

To cite this article: Nicholas Lord & Michael Levi (2024) Enhancing corporate accountability through covert situational integrity testing (CSIT), Griffith Law Review, 33:4, 432-455, DOI: [10.1080/10383441.2024.2439366](https://doi.org/10.1080/10383441.2024.2439366)

To link to this article: <https://doi.org/10.1080/10383441.2024.2439366>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 17 Dec 2024.



[Submit your article to this journal](#)



Article views: 661





[View related articles](#)



[View Crossmark data](#)

Enhancing corporate accountability through covert situational integrity testing (CSIT)

Nicholas Lord ^a and Michael Levi ^b

^aSchool of Social Sciences, The University of Manchester, Manchester, UK; ^bSchool of Social Sciences, Cardiff University, Cardiff, UK

ABSTRACT

This article explores the potential of ‘covert situational integrity testing’ as a mechanism for assessing corporate/organisational compliance with legal rules and standards, the goal being to enhance corporate/organisational accountability. There are major challenges to holding corporations/organisations to account for non-compliance: low detection levels, incomplete understanding of the inner workings of organisations, and the modest power of current social scientific research methodologies. This article argues there is scope for methodological innovation through the use of covert situational integrity testing, a variation of mystery shopping methodologies, to address some of these gaps, with a focus not on service quality or customer satisfaction but on compliance with regulatory and legal requirements, and as a data gathering tool on secretive and difficult to access areas of business operation. We focus on two examples: AML compliance by challenger banks, and the promotion of aggressive tax avoidance schemes. Our core argument is that covert situational integrity testing provides a research mechanism through which robust and systematic observational data can be collected and scrutinised by independent, third-party assessors to understand levels of organisational/corporate compliance with legal rules and standards, and by doing so, identify critical vulnerabilities and strengths in the compliance responses of organisations and industries.

ARTICLE HISTORY

Received 12 September 2024

Accepted 2 December 2024

KEYWORDS

Compliance assessment; integrity testing; corporate (non)compliance; organisational compliance; anti-money laundering; tax (non)compliance; corporate crime

Introduction

This article explores the potential of ‘covert situational integrity testing’ (CSIT) as a mechanism for assessing corporate and organisational compliance with legal rules and standards, in order to in turn enhance corporate/organisational accountability.¹ Since the work of Sutherland (1983) and of other subsequent white-collar and corporate crime/compliance scholars, we have significant evidence of the widespread and pervasive nature of varied forms of both routine and more episodic corporate non-compliance

CONTACT Nicholas Lord  Nicholas.lord@manchester.ac.uk

¹Although our focus here is on ‘corporate accountability’ in response to the Griffith Law Review special issue call for papers on this topic, we do consider CSIT to be a ‘general accountability tool’, that could be used to assess compliance levels of a range of organisational or business entities, from sole traders to large multi-nationals.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

domestically and globally, and it has been challenging to identify how to most appropriately react to these behaviours. There is a large body of literature that examines the regulation and enforcement of corporate crimes and non-compliance, and a recurring theme is that criminal law prosecution is rare (and conviction rarer still) (Lord and Levi, 2015; Levi and Lord, 2023). Risk-based, responsive regulatory models recognise the reality of scarce resources and a lack of political will to deal with corporate crimes. One effect of these pragmatic considerations is that regulatory attention is directed away from companies that seemingly want to be compliant and cooperative, and are deemed lower risk. The outcome of this and other factors tends to be low prosecution rates and/or negotiated settlements, reproducing existing political-economic conditions and structures (but there can be genuine reasons as to why prosecution is not the best method of achieving compliance and of regulating business behaviours (eg see Braithwaite and Drahos, 2000)). But full-blown criminal law enforcement policies are not practicable in the current landscape, and this means when we discuss the enhancing of corporate accountability, we need to incorporate mechanisms beyond criminal law enforcement, such as other forms of legal and extra-legal responses, if we are to avoid perpetuating a largely symbolic focus on ‘differential’ or ‘class’ justice.

When it comes to holding corporations to account for non-compliance, there are three major challenges for (a) enforcement authorities responsible for investigating and prosecuting non-compliance, and (b) researchers seeking to better understand the nature and organisation of such non-compliance. First, most non-compliance goes undetected by public enforcement authorities, so no regulatory or legal action follows for those violations. The size of the ‘dark figure’ of unreported and unrecorded corporate non-compliant behaviours is unknown, but we do know that estimations of the scale and scope of corporate offending based on enforcement data are likely to be conservative. The features of ubiquitous surveillance that are characteristic of contemporary public space and of gated communities are largely absent from corporate settings, except for the extensive dataveillance of workers in Amazon-type distribution centres and of the work phones of financial services staff for insider trading and regulatory infractions; and the mandated transaction monitoring of bank accounts for anti-money laundering and terrorist finance purposes. Some corporate violations such as dumping sewage and contaminants occur in visible public space, but economies in monitoring may fail to pick them up; while others such as dumping toxic waste in landfill may be actively concealed. Modest regulatory resources lead to the inevitable prioritisation of certain crimes or harms above others (eg foreign bribery when compared to labour exploitation) or to target cases more likely to yield an enforcement ‘result’, meaning much illegal behaviour is not addressed. Even where cases of corporate non-compliance do come to the attention of the authorities, high levels of discretion mean no further action can be a regular outcome. Second, regulators and academics face knowledge problems when looking to understand the inner workings of the corporate entity to determine whether legal violations or non-compliance is taking place or not, as such organisations and their employees are hard to access. (Inferring intent is further dimension.) As Whyte (2022: 87) notes, ‘the edifice of the corporation acts as a black box, a relatively enclosed system of social relationships’ that in turn obscures the organisation’s workings. Some enforcement authorities, such as the UK’s Serious Fraud Office, have legal powers (Section 2[a] of the Criminal Justice Act 1987) to compel corporations to provide internal materials, or to

provide information through formal interviews as part of the evidence gathering process before or after a formal investigation of possible criminal activities has begun. Failure to comply, or giving misleading or false information, can lead to prosecution. However, such powers are reliant on actual suspicions or detection of potential criminal activity having taken place, and where corporations that have engaged in pre-planned criminal activities, or represent an entirely criminal organisation, cooperation is unlikely to follow. In September 2024, the first conviction for a failure to comply with a Section 2 notice was overturned, perhaps deterring prosecution for such ‘failures’ (or decisions not to comply) in future.² Yet even if these knowledge problems could be partially addressed through speaking with insiders, accessing leaked data, or from whistle-blower accounts, remaining problems of power, or an inability to influence from the outside how corporates operate, enables non-compliant behaviours to endure. Third, current research methodologies are modest in the extent to which they enable us to understand the extent or scope of non-compliance, and ‘guilty knowledge’ or even orders further up the corporate chain. As socio-legal scholars and social scientists, the available qualitative and quantitative methods at our disposal are insufficient for building comprehensive accounts of levels of compliance by corporations. Interviews with insiders may provide useful anecdotes and shed some light on the organisation of corporate offending, analysis of case file data on known events might inform how we understand the nature of the problem, scrutiny of official statistics might indicate patterns in enforcement, and so on: but real-time insights into *current* legal violations or non-compliance are absent. These three issues present a major research challenge for socio-legal scholars, and we need methodological innovations to assist us with understanding levels of compliance with the law.

To address this challenge, in this article we propose the use of an innovative socio-legal method, that of covert situational integrity testing. In essence, CSIT is a method for generating new data by covertly but ethically participating in and observing (online) interactions between purported/fictitious clients, and organisations or corporations that offer services or products, across a theoretically selected range of scenarios or situations, in order to assess whether the latter comply with legal rules and standards relating to their service/product provision, and to collect novel and systematic data about organisational behaviours over time. The method involves some deception (ie, covert), as the organisations will not be aware that the clients are actually researchers, but that we argue to be justifiable and ethical, whilst the clients will follow particular scripts or vignettes to test how organisations respond under different situations and conditions. This (in our view, defensible) deception does, of course, raise significant ethical considerations (eg risks of entrapment, causing unfair employee dismissal, non-disclosure of the nature of the interactions, and so on), and we address these in greater depth later. By collecting data at scale and over time on such interactions, insights can be gained into the integrity of these corporate/organisational suppliers in terms of their levels of legal compliance, and identify emerging patterns and trends in compliance. In these terms, the method collects both intensive (ie qualitative insights on individual incidences of compliance or non-compliance) and extensive (ie quantitative insights on

²2 Hare Court, ‘David Whittaker KC and Gabriele Watts act for Anna Machkevitch, daughter of ENRC Founder, who has her SFO conviction overturned’ <<https://www.2harecourt.com/2024/09/19/david-whittaker-kc-and-gabriele-watts-act-for-anna-machkevitch-daughter-of-enrc-founder-who-has-her-sfo-conviction-overturned/>> 19 September 2024.

organisational or industry levels of compliance) data to make sense of organisational integrity and compliance.

With the above in mind, this article considers the scope for methodological innovation in the form of CSIT, a variation of mystery shopping methodologies, as a means of improving these data gaps and in turn testing the integrity of corporate and organisational compliance. In this sense, we propose here the implementation of mystery shopping with a twist. That is, 1. with a focus not on service quality or customer satisfaction, but on compliance with regulatory and legal requirements, and 2. as a data (and intelligence) gathering tool on secretive and difficult to access areas of business operation (eg anti-money laundering (AML) compliance). We start by visiting the methodological literature on mystery shopping to present its original purpose and contribution, before extracting some key themes that can be useful for adapting or innovating the method to be applied to corporate and organisational crimes and non-compliance. We then propose a framework for implementation aligned with CSIT. To demonstrate this, we focus here on two examples: AML compliance by challenger banks, and the online market for aggressive tax avoidance schemes. Our core argument is that in some form, the method can be and is being implemented by varied stakeholders, including academic researchers, regulators, investigative journalists, businesses etc., to gather targeted data and intelligence on particular sectors and form part of innovative attempts to recreate, and enhance, corporate accountability. It also constitutes a more objective test of experience-based heuristics, which can be self-reinforcing if not challenged by evidence.

The mystery shopping methodology

Most academic literature on mystery shopping methodologies can be found in the business, management, marketing, and administrative sciences; there is very modest literature on the use of mystery shopping as a *sociological*, *socio-legal*, or *criminological* research method. In this section we cover the origins and key features of mystery shopping methods, before adapting and innovating for use in the sphere of accountability for corporate and organisational non-compliance.

Mystery shopping as an observational method

In essence, '[m]ystery shopping, a form of participant observation, uses researchers to act as customers or potential customers to monitor the quality of processes and procedures used in the delivery of a service' (Wilson, 1998: 414). Its purpose is threefold: '(a) to act as a diagnostic tool identifying failings and weak points in an organization's service delivery; (b) to encourage, develop, and motivate service personnel by linking performance measurement tools directly with appraisal, training, and reward mechanisms; and (c) to assess the competitiveness of an organization's service provision by benchmarking it against the offerings of others in an industry' (Wilson, 2001: 732). It is an approach predominantly associated with businesses (eg, restaurants, gyms etc) seeking to complement alternative evaluation tools, such as customer surveys, to assess satisfaction with their services or products, or to manage and measure service quality, or to assess compliance by employees by inserting mystery employees into their own organisation to observe their peers (eg, to identify theft at work) (Devi and Reddy, 2016: 12).

In other words, it originated as a tool used by businesses to assess their own customer service processes, rather than by external third-party actors to assess their experiences for other purposes, although there is now an industry of expert mystery shopping companies that can provide this service for businesses.

Mystery shopping, then, is most closely associated to what we, as socio-legal scholars and social scientists, would refer to as an observational method, and a form of participant observation in particular situations most specifically. Traditionally, when implementing observation techniques, 'the primary research instrument is the self, consciously gathering sensory data through sight, hearing, taste, smell and touch' (Jones and Somekh, 2005: 138). Gans (1968) provides a classification of participant observer roles and views them as coexisting in any research project. These are: 'total participant' – the ethnographer is completely involved and resumes the researcher position once the situation has unfolded; 'researcher participant' – the ethnographer partially participates so that he or she can function as a researcher throughout; and, 'total researcher' – the ethnographer observes without involvement and therefore has no influence on the flow of events. In an ideal world, an investigation into organisational compliance would involve a degree of sustained immersion in the form of participant observation or 'observant participation' (Nelken, 2000: 25) in the working environments of corresponding organisations and their actors but, as discussed above, this is rarely feasible due to access challenges. Access to lawyers' interactions with clients (and even interviews with lawyers about their interactions with identifiable clients) would raise severe professional and legal issues since Legal Professional Privilege (or 'Professional Secrecy' in continental European language) has special protected status (Middleton and Levi, 2015; Parker and Evans, 2018; Levi, 2022). But mystery shopping can offer a form of participation and observation of particular situational interactions at a distance in cases of corporate (non)compliance. In these terms, mystery shopping methodologies are distinct from fully immersive research methodologies, such as ethnography, where global cultural understandings are sought, rather than a focus on particular behavioural processes or outcomes.

However, observation techniques have their caveats. As Adler and Adler (1994: 381) note, there are two chief criticisms of observation. First, observation methods have problems of validity. Observers must rely on their own perceptions meaning bias from their subjective interpretations of situations is inescapably evident. As Jones and Somekh (2005: 138) argue, human behaviour is highly complex rendering it impossible to make a complete record of all the researcher's impressions. They further argue that the subjectivity of the researcher throughout the research process is extremely influential given that the recorded observations become a product of choices about what to observe and what to record. Second, observational research lacks reliability. While naturalistic observation enables insights into the group or individual observed such findings are not generalisable: for instance, the insights gained from a small number of private sector organisations do not reflect the private sector as a whole. However, Adler and Adler (1994) do suggest that observational research conducted systematically and repeatedly over varying conditions that produces the same findings can be given more credibility.

With mystery shopping methodologies, the focus is on the assessment of the process rather than the outcomes of the interaction with the business, 'looking at which activities and procedures do or do not happen rather than gathering opinions about the service experience' (Wilson, 1998: 415). 'Researchers' (ie usually regular customers) would be

recruited to implement the process, and would need recompense (usually financial) and training on how to follow particular scenarios or guides, how to record data, and so on, and these interactions would take place in-person, on the phone, or increasingly online (emails or websites). Researchers must also learn how to retain their covert status and maintain confidentiality whilst remaining ethical in their interactions. A key issue with the method is the extent to which consent is required – a business looking to assess its own processes, internally or via the use of market research companies, may make employees aware that such mystery shopping may occur at some point, but where organisations are not notified in advance of the intention to mystery shop its processes, ethical issues may arise. When organisations themselves undertake mystery shopping, research suggest employees' acceptance of this is critical and that the novelty can wear off, leaving to complacency and demotivation (Wilson, 2001). There is no 'one size fits all' approach when it comes to the technique, with great diversity in its application relating to the number of researchers recruited, the number and nature of interactions observed, the purpose or rationale for the method's use, and so on.

A systematic review of mystery shopping literature as a tool to measure public service delivery by Jacob et al (2016: 165) demonstrated how such methodologies, whilst originally the preserve of the private sector, are also now emerging within the public sector for the purpose of accountability and performance monitoring (ie evaluation of service quality and evaluation of public policy interventions). For instance, public policies or programs may be evaluated to assess levels of compliance eg to assess whether alcohol or cigarette vendors are meeting legal requirements, such as relating to age limits, or to check compliance of pharmacists with legal requirements. It may also be used by Trading Standards and suchlike bodies to test whether genuine or counterfeit tobacco, vape and other products are being sold knowingly or recklessly to under-age persons. This is a significant departure from the original mystery shopping methodologies, as in the above examples, the 'mystery shopping' would be undertaken by regulatory bodies and authorities with punitive capacities in a context of law enforcement, rather than by businesses in a context of customer-driven assessments. Drawing on this use of mystery shopping in the sphere of public services in relation to accountability and performance monitoring, we see potential for the use of the technique by public authorities to assess compliance levels of the organisations for which they have responsibility to regulate, but there are several key issues that arise.

Mystery shopping in other sectors

We have also seen anecdotal use of mystery shopping methodologies elsewhere outside of academia. Investigative journalists regularly employ a variation of the method. For instance, a 2024 BBC investigation in the UK to assess the smuggling of opioids (nitazenes) into the UK involved the journalists posing as a drug dealer and contacting 35 online suppliers of the drugs to see which suppliers would send to the UK.³ A former News Of The World investigations editor, Mazher Mahmood, became known as the 'Fake Sheikh' as he regularly posed as a sheikh to uncover unscrupulous and criminal

³BBC News Online, 'Deadly opioids smuggled into UK in dog food, BBC learns' <<https://www.bbc.co.uk/news/uk-68712372>> 22 April 2024.

behaviours, in some cases by celebrities such as sportspeople, Royal Family members, and actors – he claims to have secured 100 convictions over 20 years.⁴ There are many other such anecdotal instances of investigative journalists using questionable methods to ‘entrap’ individuals. (The editing of incidents by journalists or academics can be critiqued by the subjects of their investigations and – if not transparent – can lead to miscarriages of justice.) Mystery shopping has been used as a method of consumer protection by the Financial Conduct Authority (FCA), and its predecessor the Financial Services Authority (FSA). In describing mystery shopping, the FCA’s handbook states: ‘Representatives or appointees of the FCA (which may include individuals engaged by a market research firm) may approach a *firm*, its agents or its *appointed representatives* in the role of potential retail *consumers*’ (FCA, 2024: para 2.4.1, italics in original). The FCA goes on to explain that ‘by recording what a *firm* says in discussions with a ‘mystery shopper’, the FCA can establish a *firm*’s normal practices in a way which would not be possible by other means’ (FCA, 2024: para 2.4.2). For instance, in 2019 the FCA (FCA, 2019) published findings from its mystery shopping review of motor finance discretionary commission models and consumer credit commission disclosure. This exercise involved the FCA visiting 122 motor retailers and brokers to assess compliance levels. The FCA recognised that the sample size was small and biased towards independent retailers offering specific financial solutions (eg, PCP or hire-purchase), but they found, amongst other things, that only a small number of brokers disclosed to their customers that commissions may be received for arranging finance. Thus, mystery shopping is a mechanism for the FCA to gather information about regulated actors and their service provision.

From anecdotes to social scientific rigour

The mystery shopping methodology is not new, but it has often been employed in informal or commercial settings, with a focus on anecdote rather than systematic and robust data collection and analysis. Where studies have attempted to introduce greater rigour, aiming to obtain large-*n* samples for better representation of a phenomena and/or inform policy impacts and change in the public sector, they have often been termed field experiments, and more specifically, audit studies.⁵ ‘Audit studies generally refer to a specific type of field experiment in which a researcher randomizes one or more characteristics about individuals (real or hypothetical) and sends these individuals out into the field to test the effect of those characteristics on some outcome’ (Gaddis, 2018a: 5). The focus in this field has often been on discriminatory practices, and such studies have involved in-person and/or correspondence approaches, while a common limitation includes the documentation of particular phenomena (eg, discrimination, or compliance) rather than covering the mechanisms that drive these responses, calling for the combining of audit studies with other methodologies. For instance, of most relevance to this paper is the work of Sharman (2010) and Findley et al (2014, 2013). For instance, the objective of Sharman’s (2010) ‘audit study’ was to form anonymous corporate vehicles without having to provide proof of identity before then opening bank

⁴BBC News Online, ‘The Fake Sheikh’s most famous stings’ <<https://www.bbc.co.uk/news/uk-37555017>> 5 October 2016.

⁵For a comprehensive series of chapters on the use of in-person and correspondence audit studies in the social sciences see Gaddis (2018b) and for a history of audit studies see (Gaddis, 2018a).

accounts for those vehicles. As Sharman notes (2010: 129), '[t]he research design involved soliciting offers of anonymous corporate vehicles from 54 different corporate service providers in 22 different countries, and collating the responses to determine whether the existing legal and regulatory prohibitions on anonymous corporate vehicles actually work in practice'. Sharman received 45 responses, 17 of which offered to set up anonymous corporate vehicles. Obtaining a bank account was more challenging, as only five of the solicitations were successful in obtaining offers for corporate vehicles with associated bank accounts. In terms of the method, Sharman composed and sent emails to corporate service providers, including key variables such as the need for confidentiality and tax minimisation for an international consultancy project (all factors common in known illicit cases). Sharman compiled a list of corporate service providers from theoretically selected jurisdictions, and contacted a sample of these, using his own name, and coded responses in line with the anonymity variable.

Findley et al (2013, 2014) performed a randomised field experiment with global scale to assess how core principles of international relations theory affected the responses of incorporation services when foreigners enquire about opening new companies. In terms of the method, the authors 'adopted e-mail aliases, posed as international consultants, and requested confidential incorporation from 1,264 corporate service providers in 182 countries' (Findley et al, 2013: 659). Unlike Sharman (2010), 'mild deception' was involved in their field experiment which raises ethical issues that were cleared by their university ethics board and that are discussed in the article. Findley et al developed theoretical informed experimental 'treatments' (eg, compliance with international rules or not) where their email aliases were randomly assigned origin from within one of eight minor-power OECD countries alongside a placebo condition where no additional information was offered. Findley et al built a subject pool of incorporation services through standardised internet searches to identify websites from which they then extracted contact and covariate information. The subject pool sample was then stratified, and experimental conditions randomly assigned. Incorporation services received either the placebo email or one of three treatment emails, and responses coded. Security of the exercise was ensured through the use of specially created email accounts, mobile phone numbers from an African country, and proxy servers used to hide IP addresses. All identifying information was deleted post-coding. Findley et al (2013: 673) found that 26.2 percent of the service providers contacted and 48.9 percent of those who actually responded were willing to defy international standards in providing a shell company without requiring certified proof of the customer's identity. Thus, substantial non-compliance was identified.

An important point to note is that the Findley et al/Sharman method is not designed to test established networks of relationships but rather interactions between strangers. It thus underestimates the corrosive effects of insider trust in the 'good old boys' models which in general exclude strangers or others not part of the network. For instance, research by Lord et al (2018, 2019) built on the research of Findley et al/Sharman, to better understand the dynamics of company formation markets and their role in the organisation of illicit financial flows. They found that a stratified market exists, where at the 'economy' end of the market spectrum, access to company formation services was widely accessible to all who could contact agents online, in writing, or on the phone. But at the 'premium' end of the market, company formation relations (ie, between agents and clients) are established and maintained in closed, or invite-only, settings, that exclude

strangers not part of the network, and where trust between known associates is central. To get closer, one might have to recruit trusted insiders to make offers to see where the boundaries of acceptable illegality/illicit behaviour are. This might be close to getting them to 'wear a wire'. Or, as in their recent work (see Findley et al, 2022), one might resort to impersonation of known individuals (in their case, sanctioned individuals post the Russian invasion of Ukraine) to assess how formation agents interact in those circumstances.

In relation to our focus on compliance, one notable recent study is that of Diamantis et al (2024) who assessed the potential of what they called a 'closed book privacy audit' for detecting privacy violations by corporations without their cooperation. As the authors point out, computer scientists have 'learned to use what they could observe to infer what they could not' (p 4). For instance, it is possible to selectively feed fictitious personal data to online platforms and measure/track what happens to these data online, in terms of how user website experiences change (eg tailored content). These data insights can then be used to assess corporations' (mis)use of this personal information across the data ecosystem (eg illicit transfers of data). In these terms, there is a larger social responsibility justification for the deception of online companies by providing fictitious data in order to gain insights into misuse or compliance violations – a similar justification can be made for CSIT. Diamantis et al (2024) sought to bring together such closed book audits together with an analysis of privacy law disclosure requirements of companies. Although the focus here was on these privacy disclosures by corporations, and how they should be done in a standardised and consistent way, they implemented an experimental and automated method that used a web crawler to visit websites with their target data brokers and search and collect relevant data (CSIT could make use of automation in this way to extract formal data about each online corporation's compliance statements eg searching for contextual data on 'Know Your Customer' or 'Anti-Money Laundering' or 'Anti-Bribery and Corruption', and so on.). (Given the recent explosion of assistive and generative artificial intelligence (AI) and machine learning (ML), there is great potential for these tools to form part of the CSIT methodology. For instance, real-time interactions using artificially 'intelligent' actors to interact with online systems alongside enhanced automation for wider auditing potential could substantially upgrade the method.) Thus, CSIT brings similarly brings together covert, closed book audits, alongside legal compliance requirements to assess corporate compliance integrity.

Key reference groups

Mystery shopping methodologies can be useful for an array of stakeholders keen to better understand, anecdotally, systematically, or tactically, processes and outcomes associated with myriad business practices and customer relations. Academic researchers, law enforcement authorities and regulatory agencies, public and private sector organisations, investigative journalists, and so on, have gained value from making use of the methodology. Above, we have highlighted some important examples from academia (eg, Sharman, 2010; Findley et al, 2013; Diamantis et al, 2024) where the intention was to undertake systematic field studies/experiments to understand patterns of non-compliant organisational behaviours within certain sectors. In this article we promote this line of inquiry also, seeing great value in the methodology for enhancing corporate

accountability. But we recognise also that the method has been, and could be used, by other societal stakeholders, although the underlying purpose or intentions of these groups may or may not always align. For instance, investigative journalists may not seek systematic rigour and data collection to build theories of non-compliance, but seek to draw attention to a social bad, or find a newsworthy ‘scoop’ or story for their news outlets. Law enforcement or regulatory agencies may aim to reactively address complaints in particular sectors or businesses in relation to bad practice and consumer harm, or gain advantage tactically and operationally against suspicious businesses or groups. Businesses and corporations may look to assess their own internal practices without the knowledge of their employees to understand whether they are complying with regulations and standards. We see CSIT as a method that could be used by all such groups but recognise that how academics use CSIT will likely be very different to its use as an enforcement technique, albeit with shared components.

Key issues to consider with mystery shopping methodologies

There are various questions that arise when seeking to develop an appropriate implementation framework for mystery shopping. The implementation of mystery shopping can:

- be intensive (eg, with a qualitative focus seeking to understand how a process works, or what agents actually do) or extensive (eg, with a quantitative focus seeking to identify regularities, patterns or distinguishing features of a population) with corresponding data analysis approaches
- be more structured (eg, using a very clearly defined protocol for data collection) or unstructured (eg with looser parameters in terms of which data will be sought) and this will have implications for the coding scheme that is developed
- be covert (ie, implemented without the organisations or actors being aware they are being observed) or overt (eg, making organisations aware that they will be observed, or co-opting them as research partners): raising important issues relating to informed consent
- take place in different settings, such as physically (in-person at an organisation’s premises) or remotely (eg, on the phone or online via email or web cats, though the latter is less frequent)
- be undertaken at a snapshot in time for real-time insights or longitudinally to collate repeat observations over time on specified variables
- use real researcher data as part of the process (ie, real names etc) or make use of aliases to protect researcher identities; we favour the use of aliases and consider this to be theoretically and ethically justifiable.
- undertaken by researchers rather than paid participants, as is more traditionally commonplace.

In all cases, those with responsibility for the implementation of the method must be trained in doing so and must properly instruct those conducting the research, as well as test the practices of the researchers to assess whether they implement the method with rigour.

From a socio-legal or social scientific perspective, there are questions about whether the data generated can be truly indicative of organisational or business processes and

therefore legal compliance. For instance, any given interaction is context specific – an employee may have a bad day or make an error, rather than be involved in systematic forms of non-compliance. Questions arise relating to the external and internal validity of the methodology. That said, the method avoids issues relating to of behaviour change when respondents know they are being observed, or the self-selection of respondents who volunteer. Aiming for larger systematic and robust samples, rather than anecdotal insights, will enhance the rigour of what can be inferred from the collected data, reduce the bias introduced by individuals having ‘bad days’ and provide insight into levels and patterns of compliance at the organisational, and not just individual level. (Although our focus here is on corporate/organisational accountability, individual behaviours alone can also be of interest.)

As mentioned above, the issue of informed consent is significant given that its absence when aiming to test the integrity of compliance systems represents a covert form of research. In relation to covert policing, earlier research from Levi (1995) indicated the use of covert operations in UK white-collar crime investigations has played an insubstantial role; this remains the case in 2024 and for the same reasons. As Levi noted, the reasons are i. cultural, in terms of unimaginative, conservative attitudes to detective work combined with opposition to ‘foreign’ ideas; ii. cost, since given finite resources which can be employed on known, reactive work which does not have to be specifically authorised by senior officers and whose effects are not particularly visible, covert policing is relatively risky and expensive; and iii. legal, in terms of uncertainty about the admissibility of evidence and, more probably, explicit statements of judicial disapproval of their tactics, which can harm their personal career prospects as well as the reputation of what senior police managers would now term the ‘service’ rather than the ‘force’. Covert policing has many varieties, the most extreme being the use of undercover police officers. But this is a substantially distinct context when compared to the use of CSIT, as with the former, officers who undertake covert activities (including law-breaking), are protected by various laws and guidelines. In the UK, the Covert Human Intelligence Sources (Criminal Conduct) Act 2021 sought to consolidate the existing, overlapping legal frameworks to provide for the authorisation of criminal conduct in the course of, or in connection with, the conduct of cover human intelligence sources. That said, cultural, cost-related and legal barriers, as identified by Levi, may remain obstacles to public enforcement authorities making use of CSIT and other related mystery shopping or audit mechanisms. This creates a gap into which third-party academic researchers and/or civil society groups can also offer contributions, where cultural and cost-barriers can be overcome, providing the undertaking of CSIT remains legal.

There is a notable literature on the concept of deception and social science research, emanating from within psychology in particular. One strand of this focus on deception as an object of study, focuses on the nature and prevalence of lying, (self-)deception and dishonesty (and their detection) in society (see Denault *et al.*, 2022 for an overview of this literature). Our concern is more with the justification of deception by researchers as part of social science research. Here, as Kimmel (2012: 401) notes, ‘[d]eception largely emerged as a practical solution to the experimenter–participant artifact problem – the recognition that participants come to the research setting not as passive automatons who respond mechanistically to the manipulations to which they are subjected, but as conscious, active problem solvers who often attempt to guess the investigator’s

hypotheses to do the right (or good) thing'. In these terms, deception is justified on methodological grounds, as without it, many research investigations would not be able to take place. Deception can be active or passive: the former relates to the blatant misleading of participants about some aspects of the investigation; the latter relates to omission, whereby the researcher intentionally withholds relevant information from the participant (Kimmel, 2012: 402). Varied forms of deception can take place at various stages of the research, so deception is a spectrum. Furthermore, Athanassoulis and Wilson (2009) distinguish between normative and non-normative accounts of deception, with the key issue relating to whether or not a moral judgment of the deception is necessary. In other words, the presence of deception is in itself not morally problematic, but whether it is ethically justifiable is the key consideration. Arguments have been made that the use of deception in research can 'pollute' future field studies as participants may be suspicious or behave differently, but evidence suggests this 'public good' argument is not supported (Krasnow et al., 2020). Thus, the use of deception in research needs an evaluation of trade-offs alongside the nature of the research question (Eckerd *et al.*, 2020).

In terms of research investigation rather than law enforcement investigation, we might say that mystery shopping represents a form of 'active deception' in that the organisations we interact with may not be aware of the nature of the interaction (unless they participate as co-producers of the research and so approve the deception), as opposed to 'passive deception' where we may surveil without direct contact. In this scenario we must be wary of provocation or entrapment, but we consider this form of deception as mild, even if intentional. The key question (for ethics committees also), is whether the withholding of information is reasonable in the circumstances and context of the research objectives (Athanassoulis and Wilson, 2009). As Findley et al, (2013: 668) justify in terms of the need for deception in their field experiments, '[a]cceptable standards for deception in the social sciences require that the benefits of the research be significant, that the costs be minimal, that the research avoids any physical or emotional pain, and that the research cannot be carried out in another way'. We believe the use of CSIT meets these criteria (though this will be tested when ethical approval is sought) but we recognise that its use must always be legal.

In the legal use of CSIT, three main issues are relevant: the potential for entrapment; the risks of unfair dismissal of employees; and the risks of researchers encountering criminal behaviours. First, as indicated in the judgment *R v Looseley* [2001] 1 WLR 2060 in England, '[e]ntrapment occurs when an agent of the state – usually a law enforcement officer or a controlled informer – causes someone to commit an offence in order that he should be prosecuted'. Thus, entrapment relates to how enforcement authorities and actors might induce *criminal* behaviours by individuals that they would otherwise not have committed. The question here, therefore, is whether CSIT could coerce or induce individual employees to engage in a criminal violation for the purposes of prosecution. The jurisdictions in which CSIT is used are relevant here.

In the UK, entrapment is not a legal defence, but related evidence might be excluded from a case if it can be demonstrated that law enforcement unduly induced a crime. In Australia, entrapment is not a standalone legal defence, but (as in the UK) there is discretion for judges to exclude related evidence if it was deemed to have been gathered

unfairly, unlawfully or improperly. In the US, entrapment can be a legal defence where the inducement can be proven and where the offender's prior predisposition to committing the violation is not provable, but it is ultimately a matter for juries to determine this. In our view, the use of CSIT does not constitute a form of entrapment, as it is a mechanism that does not *induce* certain criminal behaviours, but rather presents an opportunity for individuals to respond in a manner they choose in the course of their occupational duties, whether or not this is in line with expected/formally approved standards and regulations. Also, the intention of CSIT is not to lead to an outcome of prosecution, but to act as an informal regulatory mechanism for improving corporate and organisational compliance practices consistent with the law.

Second is the matter of employment law and unfair dismissals, and specifically concerns about individual accountability and potential scapegoating. We envisage CSIT to be used to hold to account corporations and organisational entities, rather than specific employees within those organisations, although we recognise the risk that organisations could nonetheless scapegoat particular employees should their identities come to the attention of the organisation. For this reason, we favour the use of anonymised and pseudonymised qualitative and quantitative data collected as part of the CSIT process (eg removal of any names, dates/times of interaction, and so on) to minimise the likelihood of individual employees being targeted by senior managers, and in turn shifting accountability away from inadequacies in organisational processes and structures towards individual failures. By using aggregate data about compliance levels rather than detailing identifiable information generated through specific employee interactions, we aim to avoid collateral effects relating to workplace dismissals.⁶

Third is the issue of researchers encountering criminal or unlawful behaviour. The key issue here is whether researchers must or should report this. Any such decision depends on the jurisdiction in which the researcher is based. For instance, in the UK, researchers are only legally obliged to notify the authorities when acts of terrorism, child sexual abuse and exploitation, and money laundering have taken or could take place. With CSIT, focusing on compliance with, for instance, FATF standards could relate to concerns over money laundering. If an organisation, for instance, is willing to open a bank account for a client whom they suspect of being engaged in crime or acting for an offender, this could generate risks of charges for money laundering. However, as the clients are fictitious in CSIT, no money laundering will take place. (Though this might not stop action by the Solicitors Regulation Authority, which commonly charges law firms for procedural violations even though no money laundering has occurred). Instead, the focus is on compliance with rules and standards and not on actual criminal behaviours. In such cases, we do not envisage a legal requirement to notify the authorities of particular non-compliant acts, but there may be a normative argument to do so. This will depend on the individual researcher. That said, should a researcher choose not to notify the authorities, they must consider the data security and anonymisation practices that accompany

⁶Though if there are regulatory, civil or criminal lawsuits, issues of disclosure to defence or prosecution may arise which may be hard to resist. See the long saga over the Boston University oral history project containing interviews with senior IRA members, which the US courts required to be handed over to the Police Service of Northern Ireland for evidence, a case that involved the highest courts in the US and UK, and not finalised in the UK courts as we write more than a decade later (Breen-Smyth, 2019; Kara, 2022).

their research, as they may at some point be compelled to provide such data to the authorities. However, sharing insights in the aggregate of compliance levels across sectors and industries should be encouraged.

A final ethical issue relates to the intrinsic (ie decisions of the researchers about the research process) and extrinsic (eg interests of institutions and funders) politics of the use of CSIT. In intrinsic terms, the choice to focus on specific sectors and specific types of compliance obligations reflects the interests and values of the researchers. For instance, by focusing on one particular group ahead of another, such as new FinTech companies rather than established 'Big Tech' companies, the research might subsequently harm the development of disruptive new businesses that offer greater accessibility to financial or technological services, or reinforce the market dominance of enduring large organisations. Relatedly, the sources of funding or the institutional context of the researchers (eg funding from, or researchers working within a mainstream bank) are significant, and can lead to extrinsic influences over the direction of the research. Navigating these political issues can be challenging for researchers who must take informed decisions about the focus of their study. These intrinsic and extrinsic politics are evident in any research project and self-reflection in the decision-making process is necessary. The concerns of civil society and/or policymakers can usefully direct our research focus, as can systematic, evidence-based analyses of the harms of different sectors and non-compliance behaviours.

A framework for covert situational integrity testing (CSIT)

Most corporate and organisational crimes and non-compliance are investigated reactively by enforcement and/or regulatory authorities, if at all, and most covert aspects of this relate to the obtaining of financial information, for instance, about suspected bank accounts without the knowledge of the offenders. Proactive covert investigation of these offenders is rare in the UK, but to a greater extent elsewhere such as the US, whilst covert social scientific research has ethical implications. With the above discussion of mystery shopping methodologies in mind, we propose here the implementation of a variation of mystery shopping, that of CSIT, as a mechanism for integrity testing with a twist. That is, 1. with a focus not on service quality or customer satisfaction, but on compliance with regulatory and legal requirements, and 2. as a data gathering tool on secretive and difficult to access areas of business operation (eg company formation). There are likely to be more advantages of mystery shopping methodology also. In terms of 1., CSIT can be seen as a mechanism for identifying points of vulnerability and/or points of strength in an organisation's processes (on a spectrum). In terms of 2., the mechanisms can encourage and develop real-time datasets on compliance and/or other market significant aspects. [Table 1](#) outlines the key features of, and the main similarities and differences between, earlier iterations of mystery shopping methodologies, as discussed in the review of literature above, and our proposed CSIT model.

There are notable analytical similarities between approaches generally aligned with mystery shopping and our proposed compliance-oriented CSIT model. For instance, both represent forms of participant observation with researchers covertly posing as potential customers in order to access the natural environments and situations, and

Table 1. Comparing mystery shopping with CSIT.

Mystery shopping methodologies	Covert situational integrity testing
SIMILARITIES	
<ul style="list-style-type: none">• A form of participant observation of situational interactions• Uses researchers to act as (potential) customers• Covert in nature so that interactions are natural• Involves deception	<ul style="list-style-type: none">• A form of participant observation of situational interactions• Uses researchers to act as (potential) customers• Covert in nature so that interactions are natural• Involves deception
DIFFERENCES	
<ul style="list-style-type: none">• A primary focus on customer service experiences (quality and satisfaction)• Assessment of process• The individual as the unit of analysis• Concerned mainly with the behaviours and compliance of individual employees• Traditionally undertaken by businesses themselves, but increasingly by public entities• Traditionally used by internal assessors on behalf of the business• More aligned to single case studies• A focus on anecdote• Usually more immersive with observation in-person• Rarely, if ever, automated leading to small-n data collection• Generates mostly qualitative data• Ethical issues less significant if employees consent to being ‘mystery shopped’	<ul style="list-style-type: none">• A primary focus on compliance with legal/regulatory rules and standards (soft and hard law obligations)• The corporation/organisation as the unit of analysis• Concerned mainly with the behaviours and compliance of organisations• Can be undertaken on, with, or by any public or private service provider• Intended for use by external assessors of the business• Assessment of process (behaviours) and outcome (integrity assessment)• A focus on systematic and robust data collection• Closely aligned to systematic field-experiments and audit studies of single and industry-wide cases: large-n data collection• Generates qualitative and quantitative data• Less immersive with observation at a distance (ie through correspondence)• Aims for automation (using AI/ML) for large-n, real-time data collection• Identifies organisational compliance vulnerabilities and strengths• Ethical issues more significant as no consent, but justifiably so.

the ways of working, of organisations. But there are also notable analytical differences. First, we shift away from customer service oriented assessments towards a focus on corporate and organisational compliance with legal rules and standards: it is the corporation/organisation, rather than individual employees, that is the CSIT unit of analysis, and though occasional observations may not be able to tell us much about the general level of non-compliance, they do tell us something about claims of total compliance. There is a broad spectrum of non-compliance in organisations, from individual ‘bad apples’ to ‘rotten/bad barrels’, where we see non-compliant behaviours that vary in terms of their levels of organisation and pervasiveness across the organisation. Traditionally, mystery shopping has been undertaken internally by businesses themselves, or assessors (eg market research companies) on behalf of the business, which in turn leads to a more single-case study type approach with the emergence of anecdotes about particular employee interactions and behaviours. On contrast, CSIT is a mechanism that can be used by both private and public service providers, though we highlight the potential for academics to act as independent third-party external assessors, examining the compliance levels of particular businesses and groups of businesses in a sector or industry more widely, and where we seek to analyse the situational interactions (processes) and the outcomes of the interactions (compliant or not). In these terms, we advocate for a less immersive approach with observation at a

distance, such as through online correspondence. Here it is desirable to pursue systematic and robust data collection (of qualitative and quantitative data), closely aligned to field experiments and audit studies. But we aim to go further and undertake real-time, automated data collection of compliance responses undertaken with artificial intelligence / machine learning capabilities to identify organisational compliance vulnerabilities and strengths. Both approaches generate ethical issues, but these are more prominent with CSIT, as it may not seek consent (but could do so if collaborating with particular organisations), provided that, as argued earlier, this lack of consent is proportionate and justifiable. Different people may take reasoned different views on proportionality and justifiability.

But what should this approach look like in practice? A useful source of guidance on the use of mystery shopping methodologies that is of more relevance to the focus here comes from the Consultative Group to Assist the Poor (CGAP). CGAP is global partnership of over 30 development organisations.⁷ It works to advance the lives of people living in poverty, especially women, through financial inclusion. CGAP promotes mystery shopping as a useful market monitoring tool (CGAP, 2022b), providing technical guides on using mystery shopping for financial services (Mazer, Gine and Martinez, 2015) and for digital financial services (Kaffenberger and Sobol, 2017), as well as case studies of where the methodology has been applied, such as in Russia where the method identified ‘noncompliance with consumer protection regulations, consumer discrimination (including gender discrimination), and abuse’ (CGAP, 2022b: 1).

CGAP (2022a: 1) states that:

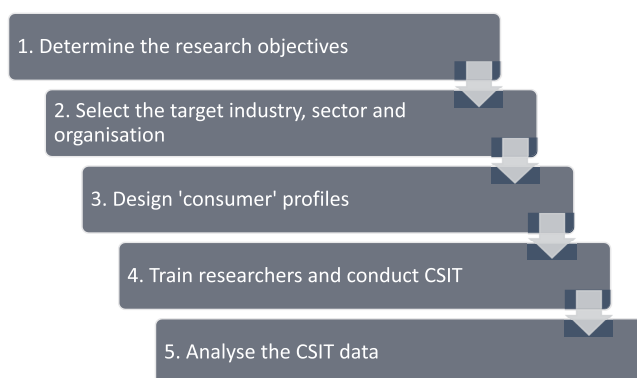
[m]ystery shopping aims to observe the actual behaviour of individual financial services provider (FSP) staff members or of third parties acting on their behalf during a true customer/FSP interaction. To use this tool, a market conduct supervisor (MCS) sends a trained consumer or supervisory staff member to an FSP access point to simulate a typical customer interaction. This “mystery shopper” then reports on their experience in a detailed and standardized manner. The interaction may be in person or remote (e.g. phone call, web chat inquiry) and relate to any part of the customer journey (e.g. shopping for a product, purchasing a product, making a transaction, calling customer service, making a complaint).

In terms of FSPs, the method can be used (a) to understand compliance with a regulatory regime, (b) to gauge the effect of a recent regulatory reform at the retail level (eg, before and after testing), (c) to assess staff knowledge at an FSP access point, and (d) to identify variations in staff behaviour at FSP access points based on customer profiles. Each of these uses can be relevant for corporate integrity testing. In our view, covert testing of situational interactions is justifiable and desirable.

⁷Including state authorities such as UK Aid and GIZ, as well as supranational organisations such as the European Commission, European Investment Bank, World Bank, and the United Nations Development Programme, in addition to private partners such as the Mastercard Foundation, British International Investment, and the Bill and Melinda Gates Foundation.

CSIT in practice

In terms of implementation of CSIT, we can adapt CGAP's (2022b) five key stages to consider when using the mystery shopping tool for social science research purposes:



In terms of the potential use of CSIT for corporate and organisational non-compliance research, our core objective would be to better understand, or measure, the nature and levels of legal compliance with existing or newly implemented laws and/or regulations in order to address knowledge gaps.

Compliance with anti-money laundering laws

For instance, our objective may be to assess compliance with a regulatory regime, such as in relation to financial institutions and anti-money laundering (AML) legal obligations. Our target industry here could be the financial sector and newly created institutions making use of financial technologies to disrupt the market dominance of the established banks. We use these new disruptive financial institutions as an example case study to reflect societal concerns in the counter-fraud communities about the prominence of accounts with such institutions being disproportionately prominent in fraud reports. For instance, a recent BBC Panorama investigation drew attention to data from Action Fraud, the UK's reporting centre for fraud and cybercrime, that indicated they received nearly 10,000 reports of fraud involving Revolut accounts (2000 more than Barclays, one of the largest UK banks) for 2023–24.⁸ The same investigation included data from the Payments Systems Regulator that showed that for every million (£) paid into Revolut accounts, £756 was from authorised push payment frauds (10 times more than Barclays). Concerns were raised in the investigation about the ease at which customer accounts could be added to new devices and ID checks easily circumvented. This programme led to further victims coming forward, as is a regular pattern in media coverage.

In this case, our 'consumer profiles', that is, our researchers, are determined by this particular objective and the context, with researchers simply looking to open accounts. We might also be interested in how banks' compliance checks vary when they encounter 'researchers' of different demographics, in which case we could vary the socio-economic

⁸BBC News Online, 'I lost £165k to fraud in an hour' – customers say they were let down by Revolut' <<https://www.bbc.co.uk/news/articles/cj6epzxdd77o>> 14 October 2024; BBC News Online, 'NHS consultant who lost £39k among 100 Revolut customers contacting BBC over scams' <<https://www.bbc.com/news/articles/c9wkzv1zk91o>> 19 October 2024.

background of our researchers. In any case, researchers would look to open accounts with start-up challenger banks in different jurisdictions to assess which information is requested and whether these requests are sufficient for meeting 'Know Your Customer (KYC)' and other AML requirements. Whilst the focus on account opening here necessarily involves the activities of individual bank employees, such as account managers, our focus is on organisational/bank compliance in the aggregate. Thus, where possible, the method could target multiple pathways to account opening to illuminate variation across the bank, or repeated account opening requests over time could be made to assess compliance levels temporally. By doing this, we shift the focus away from any given individual employee towards understanding compliance more generally by the organisation. In addition, in some cases, no human actor will be involved at all, as on-boarding practices are streamlined and automated by challenger banks to make the process more efficient. In such cases, it is the automated account opening mechanisms we would be assessing. The dynamics of CSIT will be determined by the organisational-specific products, services and mechanisms. Thus in 2022, Starling Bank in the UK was called out in UK Parliament in relation to the effectiveness of AML and counter-fraud measures as 15,000 new clients per month were 'onboarded' during the Covid pandemic.⁹ This was much greater than the market dominant high-street lenders who onboarded, on average, between 1500 and 8000, in turn raising concerns about whether proper due diligence checks were being undertaken. What is more, by June 2021 the bank had also distributed £1.6bn worth of Covid 'bounce back loans' that were subject to large amounts of fraud – the Department for Business, Energy and Industrial Strategy's estimate of fraudulent loans in the scheme totalled £4.9bn.¹⁰ Whilst Starling and other online challenger banks *may* have sufficient compliance checks in place, systematic data on relative staffing and competence are unavailable, and CSIT presents a framework for collecting such data. Within banks there is a strong focus on the customer journey from the customer experience angle and the ease of access to services, but the customer journey in terms of their ability to exploit products and services for criminal activities has not always been designed in, intentionally or not. There is consideration of 'risk' at product design but perhaps not real robust testing along the customer journey when more proactive testing is needed. The data generated here through CSIT could be analysed to provide indicative insight into how different banks apply KYC compliance checks in different jurisdictions and/or by consumers of different profiles. All this is relevant for contributing to understandings or measurements of compliance.

The following is an example of the type of email/letter that could be sent to these banks (adapted from Findley et al., 2013):

Dear [bank]
 I am seeking information on how to obtain an account for a newly incorporated entity and I hope that you might be able to offer what I need.
 I am a consultant, and my business associates and I live in Country A. Much of our business originates here, where we operate, but our company also grows quickly among international clients. Many of them are in your area so we have incorporated a Limited Company in Country B to facilitate this but require a bank account for this new venture.

⁹The Guardian, 'Starling Bank: questions over volume of customers taken on during Covid crisis' <<https://www.theguardian.com/business/2022/jun/18/questions-over-how-starling-bank-handed-out-15000-covid-loans-a-month>> 19 June 2022.

¹⁰National Audit Office, 'The Bounce Back Loan Scheme: an update' <<https://www.nao.org.uk/wp-content/uploads/2021/12/The-Bounce-Back-Loan-Scheme-an-update.pdf>> 3 December 2021.

We would like to know if you feel that you will be able to service us with an account? What identifying documents will you request for this transaction? We would prefer to limit disclosure as much as possible.

My Internet searches show that the international Financial Action Task Force sets standards for disclosure of identifying information when opening accounts. But I would like to avoid providing any detailed personal information if possible.

If you could answer these questions and also let us know about your prices, we very much appreciate it. Thank you for the time to address our query. Business obligations make communication difficult, so we would prefer to correspond with email.

Kind regards,
[alias]

Various ‘treatment’ versions of this letter can be produced and sent, varying the nature of the requests and the information shared, in order to ascertain how such variations may shape responses. For instance, if we decided to take a specific focus on ‘Politically Exposed Persons’, we might include information in the letter that ought to raise a red flag for the bank, such as the use of a particular surname in the alias alongside information about the role of the client (eg public official at a major state-owned enterprise). We can then evaluate the responses in line with legal and regulatory standards about how such ‘risky’ profiles ought to be dealt with, as dictated by the FATF standards and legal requirements in the jurisdiction where the organisation is based. Note that the regulators themselves could do these things, but this might constitute a parallel form of accountability to test the extent to which we might *reasonably* have confidence in the regulators and/or in the institutions’ claims about their own compliance.

Compliance with regulatory policies

Alternatively, our objective might be to try to understand the effects of regulatory changes. For instance, governments and tax authorities regularly make changes to what are considered permissible tax minimisation schemes. By example, in 2021, the UK tax authority (His Majesty’s Revenue and Customs - HMRC) introduced policy changes to target ‘persistent and determined’ promoters and enablers of tax avoidance schemes.¹¹ New powers, rules and legislation were proposed that would allow HMRC reduce aggressive tax avoidance, building on existing anti-avoidance measures: ‘They will reduce the scope for promoters to market tax avoidance schemes, disrupt their activities and do more to support customers to steer clear of and leave tax avoidance arrangements’. For instance, specific tax avoidance schemes, once they have been determined to be unlawful by HMRC, are added to a list of schemes not to be used, so in this context CSIT could target whether promoters continue to service clients with these schemes or not. Our target industry or sector in this example might be overseas promoters of tax avoidance schemes in the UK. Our ‘consumer profile’, that is, our researchers, would be characterised as individuals looking to minimise their income tax or national insurance contributions. As above, our researcher profiles could be amended based on demographics or other variables we are keen to better understand in terms of promoters’ behaviours with regards aggressive tax avoidance. For example, in March 2023, HMRC named two Belize registered companies as promoters of a UK tax avoidance scheme known as the Umbrella Remuneration Trust (URT) that aimed

¹¹Gov.UK, ‘New proposals to clamp down on promoters of tax avoidance’ <<https://www.gov.uk/government/publications/new-proposals-to-clamp-down-on-promoters-of-tax-avoidance/new-proposals-to-clamp-down-on-promoters-of-tax-avoidance#policy-objective>> 20 July 2021.

to avoid income tax and National Insurance Contributions (NICs).¹² HMRC drew attention to a newsletter published by the two companies, Buckingham Wealth Ltd and Minerva Services, that sought to discredit HMRC enquiries into this particular URT based avoidance scheme and encourage clients to recontribute to a new arrangement, the 'NOVA Trust', following a tribunal into the workings of Remuneration Trust based schemes. The purpose of this was to highlight the risks of the scheme to its clients and make the case that the newsletter was inadvertent admission that the URT scheme does not work. As Mary Aiston, HMRC's Director of Counter Avoidance, stated:

These cynically marketed tax avoidance schemes don't work in the way the promoters claim and users can end up with big tax bills. Over the last year we have published the details of 33 tax avoidance schemes and exposed the unscrupulous promoters behind them. HMRC has also published details of 11 Stop Notices issued to promoters, and is consulting on adding a criminal sanction for promoters who breach those notices.¹³

The list of HMRC published tax avoidance schemes¹⁴ and Stop Notices¹⁵ provides a useful starting point for CSIT. Emails can be sent to the promoters of these listed schemes with a view to ascertaining whether the schemes can still be accessed and used, or whether the promoters have complied with the regulatory change. Interactions with these promoters can illuminate the extent to which they are forthcoming or not about accepting new clients on these schemes, in turn providing insight into whether they aim to comply with or disregard HMRC notices and guidance.

The following is an example of the type of email/letter that could be sent to promoters (adapted from Findley et al 2013):

Dear [promoter]

I am contacting you as I would like your service in tax planning for my consulting firm. I am a resident of Country X and have been doing some international consulting for various [spelling mistake for added authenticity] companies. We are now growing to a size that makes tax planning seem like a wise option. A lot of our newer business is in Country Y.

My two associates and I are accustomed to paying Country Y income tax, but would like to {sp.} minimise this. We came across this tax minimisation scheme arranged by you from an internet search and would like to enquire if we could access this. [Optional: We are aware, however, that HMRC has issued a stop notice for this scheme].

As I am sure you understand, business confidentiality is very important to me and my associates. We desire as much confidentiality as we can. Please inform us also of what documentation and paperwork is required and how much these services will cost. How much can we expect your fees to be?

Due to numerous professional commitments, I would prefer to communicate through email. I hope to hear from you soon.

Best wishes,
[alias]

¹²Gov.UK, 'Evidence of marketing material used by tax avoidance promoters and suppliers' <<https://www.gov.uk/government/publications/named-tax-avoidance-schemes-promoters-enablers-and-suppliers/evidence-of-marketing-material-used-by-tax-avoidance-promoters-and-suppliers>> 14 November 2024.

¹³The Chartered Institute of Payroll Professionals, 'Offshore companies named as tax avoidance promoters' <<https://www.cipp.org.uk/resources/news/offshore-companies-named-tax-avoidance-promoters.html#:~:text=Over%20the%20last%20year%20we,promoters%20who%20breach%20those%20notices>> 10 May 2023.

¹⁴Gov.UK, 'Current list of named tax avoidance schemes, promoters, enablers and suppliers' <<https://www.gov.uk/government/publications/named-tax-avoidance-schemes-promoters-enablers-and-suppliers/current-list-of-named-tax-avoidance-schemes-promoters-enablers-and-suppliers#list-of-named-tax-avoidance-schemes-promoters-enablers-and-supplier>> 14 November 2024.

¹⁵Gov.UK, 'List of tax avoidance schemes subject to a stop notice', <<https://www.gov.uk/government/publications/named-tax-avoidance-schemes-promoters-enablers-and-suppliers/list-of-tax-avoidance-schemes-subject-to-a-stop-notice>> 14 November 2024.

As with the AML example, various ‘treatment’ versions of this letter can be produced and sent, varying the nature of the requests and the information shared, in order to ascertain how such variations may shape responses. It might be that real scheme names are taken from HMRC’s list and mentioned explicitly to gauge the response of the promoters.

Rich qualitative and quantitative data will be generated through CSIT that is suitable for intensive and extensive analysis. For instance, textual responses from organisations will enable discourse and thematic analysis. Analysing the language used in terms of how organisations respond to specific requests will provide insights into their intentions. Analysing the themes raised in responses will provide useful insights into their compliance mindsets. Numeric data can also be generated as responses are coded based on whether they comply or not with the legal requirements relating to due diligence. For instance, if a particular sector in relation to tax avoidance schemes is targeted, we can code the responses of each promoter in line with the variables included in the correspondence letters. These coded, numeric data can then be used for statistical analyses in line with the various ‘treatments’ of the correspondence submitted to organisations allowing us to understand, first, in descriptive terms, what proportions of particular organisations in certain sectors respond in the ways that they do, and second, whether there are patterns or regularities in relation to how the responses correlate with different treatments of each variable. This, of course, would require a large-n sample to be collected.

Concluding thoughts

In this article we have demonstrated the potential for a variation of mystery shopping as a tool or method for testing the integrity of corporate and organisational compliance systems and to collect data on levels of compliance, that of covert situational integrity testing. CSIT should not be a sole determinant of compliance assessments but can contribute to such assessments alongside other data. There are opportunities for testing appropriate advice and gain insights into levels of compliance with domestic and international standards, rules and laws, but great thought is needed with regards the upfront costs. There are parallels between this approach and what we see in terms of penetration testing by internal and external cyber security teams that aim to breach security measures and hack into systems in order to identify system vulnerabilities – so-called red team vs blue team type testing. This of course can also be beneficial for organisations looking to improve their compliance and the public legitimacy of their compliance claims, as well as for social science researchers. However, there are of course limitations to CSIT. While the method offers the potential to gauge baseline compliance, it cannot account for insider threats – such as complicit account managers – or repeat trust relationships in non-compliance, as this would require a deeper level of investigation including access to internal systems, policies, and audits, and so on.

There are notable ethical issues raised through CSIT, but we agree with Findley et al (2013, 2014) that mild deception is justifiable theoretically, methodologically and ethically. We agree with CGAP (2022b: 2), that mystery shopping can be beneficial in the following ways: comprehensiveness (identifying, confirming, and acquiring in-depth knowledge and understanding); proactivity (uncovering how those targeted respond in particular situations/scenarios); supervisory effectiveness (strengthening supervisory

activities); segmentation (gathering anecdotal evidence on variations in the experiences of different groups); feedback (providing input into the design and amending of regulations); and, dissemination (providing insights into key issues). In addition, the method provides potential for cross-sectional, longitudinal, case study and comparative research designs that can be automated through machine learning for real-time analysis, and also inform insights into variations across jurisdictions or sectors, and both small and big shifts in compliance levels.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Nicholas Lord is Professor of Criminology in the School of Social Sciences at The University of Manchester and the Director of the Centre for Digital Trust and Society. His research contributions centre on three inter-related areas of scholarship: (Inter)national corruption and fraud, and their regulation by criminal justice and state regulatory systems, as well as potential victims and third-party actors (business and individuals); Exploring empirically and conceptually the organisation of serious crimes for gain, in particular ‘white-collar crimes’, ‘organised crimes’, illicit financial flows and money laundering, as well as their digital underpinnings; and, the transfer of social scientific modes of analysis and thinking into the operational responses of fraud enforcement authorities. He teaches in the areas of white-collar and corporate crimes, financial and economic crimes, business compliance and regulation, serious and organised crimes, and criminological research.

Michael Levi is Professor of Criminology in the School of Social Sciences at Cardiff University. He has an international reputation for excellence in both fundamental and policy-oriented research on money laundering, corruption, cybercrimes, fraud, transnational organised crime and white-collar crimes. This is reflected in recent major lifetime achievement awards from the British and American Societies of Criminology, the Tackling Economic Crime Award, and the Al Thani Rule of Law Committee/UNODC Corruption Research and Education prize. He has played an advisory role both internationally (with the European Commission and Parliament, Europol, Council of Europe, UN and World Economic Forum) and nationally (with the UK Home Office and Cabinet Office, and with the Crime Statistics Advisory Committee).

ORCID

Nicholas Lord  <http://orcid.org/0000-0002-5922-707X>

Michael Levi  <http://orcid.org/0000-0003-2131-2882>

Reference List

Primary Legal Sources

- Covert Human Intelligence Sources (Criminal Conduct) Act* 2021 (UK).
Criminal Justice Act 1987 (UK).
Regina v Loosely [2001] 1 WLR 2060.

Secondary Sources

Articles

- Nafsika Athanassoulis and James Wilson (2009) 'When is deception in research ethical?' 4(1) *Clinical Ethics*.
- Marie Breen-Smyth (2019) 'Interviewing combatants: lessons from the Boston College Case' 15(2) *Journal of the Academy of Social Sciences*.
- Vincent Denault, et al. (2022) 'On deception and lying: An overview of over 100 years of social science research' 36(4) *Applied Cognitive Psychology* 805–819.
- S Suneetha Devi and Vidhyadhara Reddy (2016) 'A Conceptual Study of Mystery Shopping as an Ancillary Method for Customer Surveys' 16(E2) *Global Journal of Management and Business Research* 11–17.
- Mihailis Diamantis, et al. (2024) 'Forms of Disclosure: The Path to Automating Closed Book Privacy Audits' 37 *Harvard Journal of Law and Technology* 1265.
- Stephanie Eckerdt, et al. (2020) 'On making experimental design choices: Discussions on the use and challenges of demand effects, incentives, deception, samples, and vignettes' 67(2) *Journal of Operations Management* 261–275.
- Michael G Findley, et al. (2013) 'Using Field Experiments in International Relations: A Randomized Study of Anonymous Incorporation' 67(4) *International Organization* 657–693.
- Steve Jacob, et al. (2016) 'The mystery shopper: a tool to measure public service delivery?' 84(1) *International Review of Administrative Sciences* 16–184.
- Max M Krasnow, et al. (2020) 'The importance of being honest? Evidence that deception may not pollute social science subject pools after all' 52(3) *Behavior Research Methods* 1175–1188.
- Michael Levi (2022) 'Lawyers as Money Laundering Enablers? An evolving and contentious relationship' 23(2) *Global Crime* 126–147.
- Nicholas Lord (2019) 'Other People's Dirty Money: Professional Intermediaries, Market Dynamics and the Finances of White-Collar, Corporate and Organised Crime' 59(5) *British Journal of Criminology*.
- Nicholas Lord, et al. (2018) 'Organising the Monies of Corporate Financial Crime via Organisational Structures: Ostensible Legitimacy, Effective Anonymity and Third-Party Facilitation' 8(2) *Administrative Sciences*.
- David Middleton and Michael Levi (2015) 'Let Sleeping Lawyers Lie: Organized Crime, Lawyers and the Regulation of Legal Services' 55(4) *British Journal of Criminology* 647–668.
- J.C. Sharman (2010) 'Shopping for Anonymous Shell Companies: An Audit Study of Anonymity and Crime in the International Financial System' 24(4) *Journal of Economic Perspectives* 127–140.
- Alan Wilson (1998) 'The role of mystery shopping in the measurement of service performance' 8(6) *Managing Service Quality: An International Journal* 414–420.
- Alan M Wilson (2001) 'Mystery shopping: Using deception to measure service performance' 18(7) *Psychology & Marketing*.

Books

- P.A. Adler and P. Adler (1994) 'Observational Techniques' in N.K. Denzin & Y.S. Lincoln (eds) *Handbook of Qualitative Research*, Sage.
- John Braithwaite and Peter Drahos (2000) *Global Business Regulation*, Cambridge University Press.
- Michael G Findley, et al. (2014) *Global shell games: Experiments in transnational relations, crime, and terrorism*, Cambridge University Press.
- S.M. Gaddis (2018a) 'An Introduction to Audit Studies in the Social Sciences' in S.M. Gaddis (eds) *Audit Studies: Behind the Scenes with Theory, Method and Nuance*, Springer Cham, pp. 3–44.
- S.M. Gaddis (ed) (2018b) *Audit Studies: Behind the Scenes with Theory, Method and Nuance*, Springer Cham.

- Herbert J Gans (1968) 'The Participant-Observers as Human Being: Observations on the Personal Aspects of Field Work' in H.S. Becker (eds) *Institutions and the Person*, Aldine.
- L. Jones and B. Somekh (2005) 'Observation' in B. Somekh & C. Lewin (eds) *Research Methods in the Social Sciences*, SAGE Publications.
- Michelle Kaffenberger and Danielle Sobol (2017) 'Mystery Shopping for Digital Financial Services: A Toolkit', CGAP working paper prepared for the United Nation's International Telecommunication Union's Focus Group on Digital Financial Inclusion.
- Helen Kara (2022) 'Ethics Versus the Law: The Case of the Belfast Project' in D. O'Mathúna & R. Iphofen (eds) *Ethics, Integrity and Policymaking*, Research Ethics Forum 9, Springer.
- A.J. Kimmel, et al. (2012) 'Deception in research' in S.J. Knapp (eds) *APA handbook of ethics in psychology, Vol 2: Practice, teaching, and research*, American Psychological Association, pp. 401–421.
- Michael Levi (1995) 'Covert Policing and the Investigation of "Organised Fraud": The English Experience in International Context' in C. Fijnaut & G. Marx (eds) *Police Surveillance in Comparative Perspective*, Kluwer.
- Michael Levi, Nicholas Lord, et al. (2023) 'White-Collar and Corporate Crimes' in Alison Liebling (eds) *Oxford Handbook of Criminology*, Oxford University Press.
- Nicholas Lord, et al. (2015) 'Determining the adequate enforcement of corporate and white-collar crimes in Europe' in Judith van Erp (eds) *The Routledge Handbook on White-collar and Corporate Crime in Europe*, Routledge.
- David Nelken (2000) 'Virtually There, Researching There, Living There' in David Nelken (eds) *Contrasting Criminal Justice: getting from here to there*, Ashgate/Dartmouth.
- Christine Parker and Adrian Evans (2018) *Inside Lawyers' Ethics*, Oxford University Press.
- Edwin H Sutherland (1983) *White Collar Crime: The Uncut Version*, Yale University Press.
- David Whyte, et al. (2022) 'Follow the Money: Inside the Black Box of the Corporation' in M. Mair (eds) *Investigative Methods: An NCRM Innovation Collection*.

Reports

- CGAP (2022a) 'COUNTRY CASE: Russia: International Confederation of Consumer Societies (KonfOP)'. Available at: https://www.cgap.org/sites/default/files/research_documents/2022_02_MMT_Russia.pdf.
- CGAP (2022b) 'Tool 5 Mystery Shopping'. Available at: https://www.cgap.org/sites/default/files/research_documents/2022_02_MMT_5_Mystery_Shopping.pdf.
- Financial Conduct Authority (2019) *Our work on motor finance – final findings*, Financial Conduct Authority.
- Financial Conduct Authority (2024) *FCA Handbook*. Financial Conduct Authority. <https://www.handbook.fca.org.uk/handbook>.
- Michael G Findley, et al. (2022) 'Testing the Effectiveness of Targeted Financial Sanctions on Russia: Law or War?', paper presented at the 2023 Anti-Money Laundering Conference sponsored by the Central Bank of the Bahamas.
- Rafe Mazer, et al. (2015) 'Mystery Shopping for Financial Services: What Do Providers Tell, and Not Tell, Customers about Financial Products? A Technical Guide', a technical guide prepared for the CGAP.