Journal of Business Research 162 (2023) 113820

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

The anti-money laundering risk assessment: A probabilistic approach

Henry Ogbeide [a,*], Mary Elizabeth Thomson [a], Mustafa Sinan Gonul [a], Andrew Castairs Pollock [b], Sanjay Bhowmick [c], Abdullahi Usman Bello [a]

[a] Newcastle Business School, Northumbria University, Newcastle NE1 8ST, United Kingdom
[b] Independent Statistical Analyst, United Kingdom
[c] Economics Faculty, Northeastern University London, Devon House, 58 St Katharine's Way, London, E1W 1LP, UK

ARTICLE INFO

Keywords:
Financial institutions
Anti- money laundering
Expertise
Probability judgment
Bias
Overconfidence

ABSTRACT

Since the 1980s, researchers and practitioners examining the vulnerabilities of financial institutions to money laundering risk have offered some insights on how experts conduct anti-money laundering risk assessments. A common theme in the risk assessment literature is the emphasis on box-ticking rather than exercising judgment case-by-case, which has influenced our consideration of whether experts in this domain are immune to cognitive biases that novices can be vulnerable to during risk assessment. We found that both experts and novices were overconfident about their distribution judgments and this effect was slightly more pronounced in the expert group. One manifestation of the overconfidence effect in both groups was the preference for false-positive over false-negative errors. Notably, novice participants slightly outperformed expert participants in the proportion of correct outcomes. A feedback mechanism that is effective at alleviating biases, improving processes, and resultant judgment accuracy may be valuable to experts in this domain.

1. Introduction

A money laundering scheme is an action aimed at concealing ill-gotten gains that are intended for use and making them appear to originate from legitimate sources (Baldwin, 2003; Levi & Soudijn, 2020). It can be facilitated through a human medium such as money mules (persons who help third parties transfer funds using their own personal identities for commission — see Raza, Zhan, & Rubab, 2020) or information technology enabled channels like virtual currencies (Anichebe, 2020). Money laundering schemes can sometimes be quite simple; for instance, smurfing, also called structuring, which involves making multiple deposits below the threshold where banks must report cash transactions to regulatory authorities (Caulkins & Reuter, 2022; Linn, 2010; Schneider, 2020). The schemes can, however, be complicated, layered across several institutions and countries. Money launderers, for example, may use multi-layered corporate structures to consummate layers of transactions that allow the proceeds from illicit activities to be disguised as legitimate (Veen, Heuts, & Leertouwer, 2020).

Given that money laundering schemes offer varying degrees of anonymity to illicitly acquired proceeds (Kruisbergen, Leukfeldt, Kleemans, & Roks, 2019), both the private and public sectors are investing a great deal of resources in combating money laundering crimes. Government employees monitor tax declarations, while financial intelligence units (FIUs) receive and analyse suspicious reports from reporting entities (Lannoo & Parlour, 2021). Banks and other private entities undergo costly anti-money laundering procedures to help governments limit the facilitation of proceeds from crimes (Berg, 2020; Takáts, 2011). They are required, under laws predominantly based on the Financial Action Task Force (FATF) recommendations, to meet intricate compliance requirements, verify the identities of their customers', sources of funding, and monitor their payments (Pol, 2020). Similarly, under the proceeds of crime regulations, all reporting entities such as banks and their employees are required to report suspicious transactions (FATF, 2014).

One line of thinking is that suspicious activity reports (SARs) which form the cornerstone of the anti-money laundering (AML) framework, globally hang on the loose scales of suspicion (Sinha, 2014; Wilkes, 2020). In reality, what appears to be a straightforward decision is actually a complex scheme, historically carried out by public agencies such as prosecutors and courts, now entrusted to financial professionals. To fulfil this obligation, financial professionals must be able to discern suspicious behaviour within complex financial transactions using a procedural risk assessment framework (Cindori et al., 2013). This task presents a wide range of ethical, empirical, and practical restrictions when finding specific case of money laundering (Amicelle & Iafolla, 2018; Demetis & Angell, 2007; Naheem, 2017; Singh & Best, 2019), and is even more complicated when the client under risk assessment has no existing records relating to any known predicate offence (Naheem, 2019).

* Corresponding author.
E-mail address: henry.ogbeide@northumbria.ac.uk (H. Ogbeide).

https://doi.org/10.1016/j.jbusres.2023.113820
Received 30 August 2022; Received in revised form 14 January 2023; Accepted 1 March 2023
Available online 1 April 2023

Much of the surveillance in AML regimes is silent and unobserved (Halliday, Levi, & Reuter, 2019). For instance, automated systems analyse transactions for signs of anomalous informed by known typologies, or financial sector officials maintain a list of foreign and domestic politically exposed persons (PEPs) who will be subjected to increased scrutiny. Banks employ automated solutions to assess money laundering risks in large volumes of financial transactions, however, flagged transactions (from technology) still need to be manually examined by staff to determine whether they are suspicious or not (Demetis, 2010; Isa, Sanusi, Haniff, & Barnes, 2015). For the staff, the most challenging aspect is developing criteria to identify suspicious behaviour or transactions (Sinha, 2014). Without quality considerations in bank AML programs, criminals might be able to evade detection. The consequences of such failures have left major financial institutions with penalties and costs in the hundreds of millions (Nyreröd, Andreadakis, & Spagnolo, 2022). A common theme in the risk assessment literature is the emphasis on box-ticking, which results in a high false positive rate that undermines the AML system's efficacy, bank's reputation (dalla Pellegrina, Di Maio, Masciandaro, & Saraceno, 2022), and raises operating costs for law enforcement agencies that rely to some extent on these reports for intelligence (Amicelle & Iafolla, 2018; Takáts, 2011).

The adoption of the risk-based approach to AML, which means customers are risk rated based on variables such as geographic risk, customer risk, and product or service risk, which may increase or decrease the perceived risk posed by a particular customer or transaction is problematic (Bello & Harvey, 2017). Money laundering detection, for example, has to be based on a subjective assessment of an assessor, as there are no physical indicators to detect money laundering risk (Sinha, 2014). Yet, there have been relatively few studies on the individual-based role in assessing money laundering risk (Isa et al., 2015). Even though scholars in this field have noted and studied various risk assessment approaches, the studies are found to lack strong theoretical foundations for linking expert cognitive factors to the quality of AML risk assessment in financial institutions (Jamil, Mohd-Sanusi, Mat-Isa, & Yaacob, 2022). This research highlights some of the issues through the following three main research objectives. First, to examine the quality of AML risk assessment used in banks. Second, to provide an understanding of how likelihood judgments are formed within the context of AML risk assessment. Third, to further the work in the field of human judgment analysis.

AML experts face uncertainty at the core during the risk assessment procedure: interpretations are contextually sensitive, and conclusions are often probabilistic (Veen et al., 2020). Financial institutions can only see a fraction of the bigger, more complex picture when dealing with transactions (FATF, 2022). In fact, criminals exploit this information gap to layer illicit financial flows between financial institutions within and across jurisdictions. Thus, circumstantial evidence becomes the foundation of inference to identify whether proceeds originate from criminal activities (Bell, 2000). This raises the question of whether experts in this domain are immune to cognitive biases that novices are susceptible?

Probability judgment theory offers a chance calculus for measuring performance (Costello & Watts, 2014). Probabilistic judgment is essentially the assessment of the degree to which an assessor is certain that a statement is correct under uncertainty (Sanders & Ritzman, 1992). Hence, using actual financial transactions with money laundering conviction and non-conviction outcomes, we employed an experimental vignette-based survey to investigate the following research question: *How does the quality of AML probabilistic risk assessments made by experts compare to those of novices?*

The paper is structured as follows. The next section is an exposition of the theory guiding this research — the role of experts in AML risk assessment, the theory of probability judgment and the effects of expertise in probability judgment accuracy. Subsequently, the details of the methodological framework adopted in our empirical examination are presented, and this followed by the empirical findings. Finally, in the discussion of the findings, we consider how our work contributes to theory and practice and we identify research areas to explore further.

2. Theoretical background

2.1. Expert role in anti-money laundering risk assessment

Money laundering risk cannot be isolated but stems from various predicate crimes, such as human trafficking, small-scale tax evasion, forced labour, and weapons trafficking (Canhoto, 2021). Experts' ability to distinguish suspicious transactions from legitimate ones presents a dilemma, as it does with many high-stakes decisions (Amicelle & Iafolla, 2018; Bergström, Svedberg Helgesson, & Mörth, 2011). According to the regulation on suspicious transaction report, an expert must have reasonable grounds to suspect that there is a possibility that a transaction may involve proceeds from crimes (FATF, 2014). However, money laundering risk is elusive and subjective (Demetis, 2010). Its risk assessment is a judgment about risk that a counterparty or transaction may be associated with criminal funds (Van Duyne, Harvey, & Gelemerova, 2018). In fact, experts rely on a test of a transaction's economic rationality to determine whether it is suspicious or economically rationally explained (Axelrod, 2017).

Under the customer due diligence (CDD) legislation, financial institutions should take reasonable steps to identify and verify that their customers are whom they claim to be (McLaughlin & Pavelka, 2013). Remarkably, proponents of CDD have argued that data intelligence from CDD helps experts assess the likelihood of money laundering risk in financial transactions (FATF, 2014; Johari, Zul, Talib, & Hussin, 2020). In reality, experts struggle with a dilemma about ensuring sensitivity while being aware that they have conducted a reasonable assessment of the suspicious transaction (Maurer, 2005). They may also encounter uncertainties caused by incomplete or inaccurate information, transaction ambiguity, concealment, and inconsistencies during risk assessment. Implementing AML risk assessment without the appropriate risk judgment will have a negative impact on the filtering obligations financial institutions must meet (Zavoli & King, 2021). Hence, for experts to produce factual suspicious transactions, methodical suspicion becomes the main approach during risk assessment (Fedirko, 2021).

Paradoxically, money laundering risk assessment is not just a measurement exercise but a response to set requirements introduced by the AML regulatory regime championed by the FATF (Riccardi, Milani, & Camerini, 2019). These requirements are periodically updated, and guideline statement issued by FATF. For example, the FATF in 2013 created an AML risk assessment framework to guide AML regulated bodies during the assessment of money laundering risk (Halliday et al., 2019). According to the guideline, money laundering risks are assessed based on 'likelihoods' that proceeds from criminal activities will be laundered and associated 'consequences' if the proceeds is laundered (Savona & Riccardi, 2017). A disjunctive synthesis of both the likelihood and consequences judgments should guide the risk assessment of money laundering risk. For example, banks may report a transaction as suspicious if there is a high likelihood that the transaction is a case of money laundering occurring or if the consequences will be severe that the transaction turns out to be a case of money laundering.

Money laundering risk indicators (potential red flags) are also codified in national risk assessment frameworks as a guideline for identifying instances of money laundering (Savona & Riccardi, 2017). But guidelines do not usually translate easily into provider behaviour (Reyna & Lloyd, 2006). Within AML domain, experts must be able to identify and interpret erratic trails of crime proceeds across jurisdictions (Fedirko, 2021). Hence interpretation is unavoidable during AML risk assessment and all risk judgments, whether based on codified risk categories or a transaction's economic rationale, are points of inference. In such uncertain environments, probability judgment theories suggest expert contextual knowledge (work-related experience) is crucial for making accurate judgments, and technical expertise is of little significance (Sanders & Ritzman, 1992). In fact, studies have shown that experts use their contextual knowledge to interpret events based on cues they observe (Afflerbach, van Dun, Gimpel, Parak, & Seyfried, 2021).

It follows from heuristics theory that even experts make biased judgments in the face of uncertainty by using heuristics (Tversky & Kahneman, 1974). For example, several studies have demonstrated that calibration in probability judgment is adversely affected by overconfidence (McKenzie, Liersch, & Yaniv, 2008). Calibration is the measure used to describe the degree of consistency between allocated probabilities and actual occurrences (Lichtenstein, Fischhoff, & Phillips, 1977). Although, it is desirable that AML experts can defy the biasing influence from overconfidence, and they should be more suitable at doing so than novices in an AML context. However, no prior study in this domain has systematically examined whether this is actually the case.

Experimental manipulations that affect accuracy provide insight into the underlying mechanisms. One of several types of likelihood judgments is probability judgment (Yates et al., 1989). Most authors have noted that the use of probability judgment technique helps to simplify the study of likelihood judgment, since a probability is an expression of a purely internal state (Lichtenstein et al., 1977). The background of this area of research is discussed next.

2.2. Probability judgment accuracy

When examining the quality of judgment in an assessment or prediction context, it is not sufficient to judge the performance of the assessor solely by how often he or she is correct while neglecting the degree of confidence that he or she has in each assessment (Borracci & Arribalzaga, 2018; Olsson, 2014). People overestimate the accuracy of their judgments in light of their perceived overconfidence, according to psychological studies (Olsson, 2014). When subjective probabilities exceed proportions of correct answers, there is overconfidence (Garcia, Gomez, & Vila, 2022). Therefore, any effective appraising method should take into consideration both the actual assessments and their accompanying probabilities. Probability values are crucial for identifying underlying strengths and weakness of judgment in assessment such as poor calibration (Lichtenstein et al., 1977).

Probability judgments are well-calibrated to the extent that the probability judgment attached to various events matches the relative frequencies with which those events occur. For example, a perfectly calibrated assessor would be correct on seventy out of one hundred occasions where he or she provided a probability of 70%, and on 50% of occasions where he or she provided a probability of 50%, and so on. Those who are not well calibrated may either be underconfident or overconfident. In the underconfident assessment, the percentage of propositions that are true exceeds the assigned probability. While in cases with overconfidence, too few propositions are true. As mentioned above, the most recognised cause of poor calibration is overconfidence or overreaction (Lichtenstein, Fischhoff, & Phillips, 1982; Wallsten & Budescu, 1983). Another underlying aspect of judgment that can be identified from a probability judgment analysis is 'resolution' (Yates, 1982).

Resolution refers to the ability to discriminate between instances where an event is likely to take place from when it is not —- which is arguably a crucial skill in the present context of AML assessment. Yates (1982) suggests that poor calibration and resolution may result from inconsistent probability estimates from the assessor, which again would be a critical problem in the present context as similar instances should be evaluated in an equivalent manner. Money laundering risk assessment relies heavily on the interpretation and classification of transaction trails linking wealth sources to illicit activities. But AML risk assessment guidelines that centre on regulatory issues focusing on specific risk assessment approach, such as the risk-based guideline, are influenced by the practice of discriminating vague categories of risk and mapping them to distinct levels. Consequently, we anticipate that AML experts will set their risk assessment based on just a few key risk categories (such as designated customer risk, geographic risk, and transaction risk), as opposed to the many details that novices may add.

Various statistics for examining probability judgment accuracy have been proposed in the literature. Bayesian Networks (BNs), for example, express causal relationships between events using graphical inference and can be used both for predicting the probability of unknown variables or updating the probability of known variables based on evidence (Kabir & Papadopoulos, 2019). It follows a mathematical models of reasoning based on Bayesian inferences, a process for drawing conclusions given observed data in a way that follows probability theory (Costello & Watts, 2014). In a comprehensive review, Musharraf et al. (2013) evaluated the use of BNs to assess human error probabilities during offshore emergencies. They demonstrate that the BNs approach adequately assesses human error likelihood based on their comparative study. Similarly, the application of BNs in system safety, reliability, and risk assessment, was recently presented by Kabir and Papadopoulos (2019). Though BNs have gained popularity in risk assessment applications due to the model's flexible structure, there have been criticisms of bayesian models' estimation of likelihood functions and priors (Endress, 2013; Marcus & Davis, 2013). The Bayesian theory permits too many arbitrary alterations to likelihoods and priors. Bowers and Davis (2012) explain that this flexibility of the Bayesian theorem-based model could allow the usage of the model for explaining almost any behaviour as optimal.

The Mean Probability Scores (MPS) is another frequently used approach for studying likelihood judgment (Yates & Curley, 1985). The MPS is linked to Brier (1950) and is often referred to as the 'Brier Score'. It measures the difference between the assigned probabilities and whether or not the events transpired. The MPS statistic is a wide gauge of overall accuracy that can be broken down to reveal important underlying aspects of performance, such as calibration and resolution (Lichtenstein et al., 1982; Murphy, 1972a, 1972b; Yates, 1982). This study will adopt an approach based on Yates (1982), since the required estimation is simple probabilities, such as the probability of an event P(A), and does not involve any conditional probabilities of any form related to the Bayes' theorem. The approach and the relating statistics will be described in detail in the methodology section. However, in the meantime, research that has utilised probability judgment accuracy approaches to examine the quality of professional judgment is reviewed next.

2.3. The effects of expertise in probability judgment accuracy

Comparing experts' and novices' performance has historically been the most common method for studying expertise (Carter, Sabers, Cushing, Pinnegar, & Berliner, 1987). Findings from existing literature show inconsistent accuracy conclusions across different professional domains. In some fields, experienced professionals are found to be more accurate in their probability estimates than novices. For example, in the domain of clinical science, Benjamin, Mandel, and Kimmelman (2017) examined the extent to which expert cancer researchers were better able to accurately predict the probability of replicating significance levels and effect of sizes from specific original studies than their novice counterparts. However, despite their overall better performance, the study also noted that experts with specialised knowledge exhibited significant overconfidence in their area of expertise. Similarly, the work of Trueblood et al. (2018) on cancer image identification found experts' probability values were associated with higher degree of discriminability reflecting better resolution than that of novices.

Conversely, other studies have demonstrated 'inverted-expertise' effects whereby experts have performed worse than novices. For example, Parr, Heatherbell, and White (2002) explored the confidence and accuracy correlation for experts and novices in a wine odorants identification experiment and found a stronger association between confidence and accuracy for novices ($r= .60$) than for experts ($r= .24$) on the verbal memory tasks. While Larson and Billeter (2017) demonstrated in their study on expert judgment accuracy in rating twenty (20) vocalist's performance in a competition, that experts gave more critical ratings for low performance than novices. Other studies in financial contexts have illustrated inverted expertise effects (Muradoglu & Onkal, 1994; Yates, McDaniel, & Brown, 1991). For example, in a stock price forecasting study, Yates et al. (1991) found that the predictions of novices were better than those with professional experience. These authors accounted for these effects by pointing out that more experience within a domain can lead to a greater amount of beliefs being formed about the kinds of data that are predictive of important occurrences. But in areas of high uncertainty and less reliable feedback (such as in AML risk assessment) greater experience can result in a greater dependence on weak cues.

Cynics question whether experts in this domain possess enough objective information about money laundering activities to assess money laundering risks appropriately and argue that the act of not tracking and evaluating their performance against explicit benchmarks of accuracy and rigour is likely to be counterproductive.

3. Methodology

3.1. Vignette-based field experiment

Studying personal values and beliefs requires unobtrusive approaches since they are sensitive subjects (Poulou, 2001). The vignette method allows respondents to express their perspectives on topics they are familiar with, while remaining detached from them and protected from personal threat. This approach has the advantage of removing the need for respondents to be biased and give socially acceptable answers since they do not fear that honest responses might devalue their reputations (Alexander & Becker, 1978). As Kerlinger (1966) argued, vignettes combine a variety of expressive and objective ideas with projective methods, making them ideal for psychological and educational research.

Typically, in vignette studies, respondents are shown brief descriptions of situations or individuals (vignettes) to elicit their judgments about them (Atzmuller & Steiner, 2010). Aguinis and Bradley (2014) describe experimental vignette methodology (EVM) as the use of carefully constructed and realistic scenarios to assess dependent variables such as intentions, attitudes, and behaviours. There has been an increase in the use of vignettes embedded in surveys in many disciplines, like violence risk assessment (Murray & Thomson, 2010), marketing research and supply chain disruption (Cantor, Blackhurst, & Cortes, 2014). Vignettes are found to stimulate respondents' imaginations and engage them in the survey, as well as provide them with a way to express their thoughts on follow-up Likert-style formats and checklists (Poulou, 2001).

3.2. Vignette development and pilot study

The data used for the risk assessment vignettes were obtained from actual money laundering crime-related cases that included both money laundering and non-money laundering convictions outcomes. In this study, the authors identified factors of interest based on literature reviews and practice guidelines published by the FATF, as there were no existing vignettes in AML risk assessment literature that fit this study. During the design process, attention was taken to ensure that the volume and nature of the information contained within the narratives were similar to what financial professionals typically use to formulate their AML risk assessments. In Case 1, Case 4, Case 5, Case 6, Case 7, Case 9, Case 10, and Case 11, we utilise profiles designated as high-risk in FATF guidance document (Appendix A) for all three indicators (customer, service or product and geographical location) to allow for the analysis of the possible effect of this combined set on experts' judgments.

Importantly, the authors interviewed seven (7) AML experts within financial and non-financial institutions about their perspectives on the draft vignettes. The use of expert samples for pre-test purposes is generally recommended to establish content validity (Paddam, Barnes, & Langdon, 2010; St.Marie, Jimmerson, Perkhounkova, & Herr, 2021). The vignettes were then revised in response to feedback received from the AML practitioners, before a pilot test was conducted using six volunteer compliance officers to gauge their opinion before uploading the final version of the vignettes on a hosted web site. This action was intended to determine if the vignette presented a credible, realistic scenario to the average targeted respondent. Thus, the external validity of utilised vignettes was considered to be appropriate for their intended use.

This study used a vignette-based field experiment where varying versions of vignettes were used to depict the context and information about the risk-based approach (i.e., customer business lines, financial products and services, and domicile location) to human subjects. The vignettes were developed according to the guidelines outlined by Cantor et al. (2014), which call for a common module that provides contextual information that is intended to be invariant across a variety of versions of the vignette. The details of the vignette included participant role, common module, and experimental cues (see Appendices B and C).

3.3. Participants and procedure

The participants comprised 155 individuals from 13 countries (see Appendix D) who participated in the study, of whom 80 were experts from the commercial banking sectors and 75 were novices from postgraduate business schools with no experience in money laundering risk assessment. Participants in the expert group were financial institution employees with expertise in areas pertinent to anti-money laundering risk assessment and are responsible for AML activities in their institutions. The mean years of experience in AML functions within this group was 6.7 years.

The participants first read the experimental instructions from an invitation email and Page 3 of the online survey document reiterated these instructions. The instruction stated that the purpose of the study was to examine professional' AML risk assessment and requires each participant to assess twelve cases of customer financial transactions with money laundering and non-money laundering conviction outcomes. For each presented scenario, each participant would select either "Yes" or "No" if they thought there was any suspicious activity relating to money laundering in the case that might lead to a money laundering conviction outcome or not. They are also required to indicate their certainty in percentage confidence estimate between 50% to 100%. These scores provided the medium for comparing the performance of the participants and for conducting the probability judgment analysis.

3.4. Statistical framework for examining the quality of AML risk assessment

The participants made probability assessments on each of the 12 cases using two components. First, they stated whether they believed there was suspicious activity of money laundering with a simple yes/no answer. A yes answer was given a value of unity and a no answer a value of zero. Second, they stated how confident they were on their above answer by providing a probability, expressed as a percentage figure, between 50% and 100%. The analysis converted these values to probability terms between 0.5 and 1.

This is termed the half range method of obtaining probability assessments. More formally, the half range method requires the subject to make probability assessments involving two stages. In the first stage, the subject specifies whether they think the event is likely to occur. This can be denoted $d_{i,j}$ for event 'i' by assessor 'j', where $d_{i,j} = 1$ when the event is assessed as being likely to occur and $d_{i,j} = 0$ when the event is assessed as being not likely to occur. In the second stage, the subject specifies a probability between 0.5 and unity relating to the likelihood of the event occurring or not occurring that the subject had specified in the first stage. This assessment can be denoted $r_{i,j}$ for event 'i' by assessor 'j'.

In the evaluation of a participant's probability assessment, it is necessary to have ex-post outcomes. Regarding the money laundering data set, as there exists no exact data that could be used to generate the outcome data (the guilt or innocence of the defendant in the cases is not known with complete certainty), the obvious choice is the dichotomous trial outcome probabilities (0 for non-guilty and 1 for guilty). In the context of the Money Laundering data, N = 12, and trial result values, can be denoted e_i, (for the 12 cases the values in ascending order are 0,1,0,1,1,0,0,1,1, 0,0,1, where e_i, for $i = 1,\ldots,12$, equals zero for an actual non-guilty verdict zero and unity for an actual guilty verdict. In the case of half range probabilities, when $d_{i,j} = e_i$, (with e_i measured on a dichotomous scale), the outcome index is equal to unity, $c_{i,j} = 1$. When $d_{i,j} \neq e_i$, the outcome index is equal to zero, $c_{i,j} = 0$. Accuracy analysis involves comparing $r_{i,j}$ with $c_{i,j}$ for event i for assessor j.

3.5. The Mean Squared Probability Score (MSPS) and the mean outcome index

The mean outcome index, Mean(c_j), is a simple measure of overall accuracy for an assessor, j, used with the half range method. For dichotomous probability outcomes, it is simply the proportion of correct assessments. It can be defined as equation as:

$$Mean(c_j) = \frac{1}{N}\sum_{i=1}^{N} c_{i,j} \tag{1}$$

The value to be better than chance should be above 0.5 with the best possible value unity. Values below 0.5 are poorer than chance. The Mean Squared Probability Score (MSPS) is a quadratic loss function used to evaluate the performance or accuracy of a set of probability assessments. It is often referred to as the Brier Score using ex-post dichotomous outcomes. Therefore, it is necessary to have ex-post outcomes. The MSPS is analogous to the Mean Squared Error, and like the MSE, it can be decomposed into components involving bias and variation that can be used to consider specific aspects of performance or accuracy. The overall probability performance of a set of assessments for assessor j, can be measured by the MSPS, which is the average of the squared assessment errors, where in the case of the half range method is the assessment error, measured as the assessment probability value minus the outcome index value. The $MSPS_j$ is defined in Eq. (2):

$$MSPS_j = \frac{1}{N}\sum_{i=1}^{N} (r_{i,j} - c_{i,j})^2 \tag{2}$$

A value of zero would imply that assessment probability values are identical to the outcome index values (indicating perfect accuracy, that is, all probability assessments equal unity and have the correct outcome); hence, the higher the value of the MSPS the poorer the performance.

3.6. Specific aspects of accuracy or performance, the statistical decomposition of the MSPS

As discussed previously, the MSPS is an overall performance measure which can be decomposed to identify specific components that reflect the multidimensional aspects of accuracy. Expanding Eq. (2) gives Eq. (3):

$$MSPS_j = Var(r_j) - 2Cov(r_i, c_j) + [Mean(r_j) - Mean(c_j)]^2 \tag{3}$$

Where,

$Mean(r_j) = \frac{1}{N}\sum_{i=1}^{N} r_{i,j}$

$Var(r_j) = (\frac{1}{N}\sum_{j=1}^{N} {r_{i,j}}^2) - Mean(r_j)^2$

$Var(c_j) = (\frac{1}{N}\sum_{j=1}^{N} {c_{i,j}}^2) - Mean(c_j)^2$

$Cov(r_i, c_j) = (\frac{1}{N}\sum_{j=1}^{N} c_{i,j}\, c_{i,j}) - Mean(r_j) * Mean(c_j)$

Given the bivariate linear regression equation of $r_{i,j}$ on $c_{i,j}$ of form: $r_{i,j} = K_j + (SL_j * c_{i,j}) + u_{i,j}$ where,

K_j is a constant coefficient

SL_j is slope coefficient which can be considered a measure of resolution,

$SL_j = \frac{Cov(r_i, c_j)}{Var(c_j)}$

$u_{i,j}$ is an error term

Taking the variances gives:

$$Var(r_j) = [{SL_j}^2 * Var(c_j)] + Var(u_j) \tag{4}$$

Substituting Eq. (4) into Eq. (3) gives Eq. (5):

$$MSPS_j = [{SL_j}^2 * Var(c_j)] + Var(u_j) - 2Cov(r_i, c_j) + Var(c_j) + [Mean(r_j) - Mean(c_j)]^2 \tag{5}$$

Where, ${SL_j}^2 * Var(c_j)$ is the minimum variance of r (Yates)

$Var(u_j)$ is scatter (Yates) or error variation

$Var(u_j) = Var(r_j) - [{SL_j}^2 * Var(c_j)]$

$[Mean(r_j) - Mean(c_j)]^2$ is bias squared

Eq. (5) is essentially, the decomposition of the MSPS used by Yates. Given

$Cov(r_i, c_j) = 2 * SL_j * Var(c_j)$

Results in the MSPS decomposition used in Eq. (6):

$$MSPS_j = [Mean(r_j) - Mean(c_j)]^2 + [(1 - SL_j)^2 * Var(c_j)] + Var(u_j) \tag{6}$$

This decomposition presented in Eq. (6) can be presented as follows:

$MSPS_j = BiasSquared_j + ResolutionVariation_j + ErrorVariation_j$

3.7. The interpretation of the components of the MSPS decomposition

The decomposition discussed involves bias squared (BS), resolution variation (RV) and error variation (EV), using the MSPS decompositions presented in Eq. (6). That is for assessor j:

$MSPS_j = BS_j + RV_j + EV_j$

These three components are discussed next in the context of the half range method using a dichotomous outcome index. When analysing judgment, it is appropriate to compare a participant's performance with that of a hypothetical random assessor and perfect assessor. The random assessor assigns all probabilities as 0.5 with an arbitrary choice. Thus, the value of Mean(c) for the random assessor is 0.5. Therefore, Mean(c_j) for assessor j, should be above 0.5 with the maximum being unity. The perfect assessor makes correct probability assessments of unity.

The Mean Response {Mean(r)} is the mean of the $r_{i,j}$'s, viz. $\frac{\sum r_{i,j}}{N}$, where $r_{i,j}$ (which is between 0.5 and 1) is the probability response for case i, ignoring whether or not the prediction is in the correct direction. Mean(r) has, of course, a value of 0.5 for the random assessor and unity for the perfect assessor.

Bias (B) is the difference between the mean response and the mean outcome index $\{B_j = Mean(r_j) - Mean(c_j)\}$ and measures the degree of overconfidence (if positive) or underconfidence (if negative) in the assignment of probabilities and directional responses. This measure has a theoretical value of zero for the random assessor. Bias Squared (BS) is simply the square of the bias term and is a component of the MSPS. The best value is zero, which would be the case for the random and perfect assessor.

Slope (SL) or resolution is a very important aspect of performance that measures the degree to which higher probabilities are assigned for correctly assessed values. SL is the slope coefficient of the fitted simple linear regression of the responses ($r_{i,j}$) on the outcome index values ($c_{i,j}$). For the perfect assessor, $r_{i,j} = c_{i,j}$ for all assessors, so the closer SL_j is to unity the better the performance. It has value zero for the random assessor. SL is probably the most difficult measure on which to obtain good performance and is a component, in practice, that can often

Table 1
Outcome index for each case.

Case	Trial outcome[a]	Mean expert	Mean novice	Comparison significance (p)[b]
1	0	0.24	0.44	0.008*
2	1	0.80	0.68	0.089
3	0	0.21	0.12	0.125
4	1	0.51	0.64	0.110
5	1	0.66	0.67	0.956
6	0	0.41	0.43	0.859
7	0	0.46	0.33	0.102
8	1	0.73	0.79	0.374
9	1	0.71	0.73	0.773
10	0	0.38	0.37	0.983
11	0	0.34	0.32	0.817
12	1	0.56	0.63	0.418

[a]Definition of values (0 = not convicted, 1 = convicted).
[b]Pairwise comparisons (expert vs novice) via Mann–Whitney U test and Kruskal–Wallis Test.
*p <.05.

turn out negative, depending on the difficulty of the task. The related resolution component (RV) of the MSPS relates the slope or resolution component to the variance of the outcome index. The best value on this measure is zero.

Scatter (SC) or error variation (EV) relates to the degree of variation in the responses that are not explained by variation in the outcome index. It is essentially the residual sum of squares in the simple linear regression of the responses ($r_{i,j}$) on the outcome index values ($c_{i,j}$). This measure has a value of zero for both the random assessor and the perfect assessor.

4. Results

4.1. The mean outcome index value

The examination of the average outcome index values $Mean(c_i) = \frac{1}{j}\sum_{i=1}^{j} c_{i,j}$ (where j = 1,2,…,j and j = 80 for experts while j=75 for novices) for each case for the expert and novice participants shows some interesting results when compared with the actual case results. Table 1 indicates that the values are below 0.5 for the non-guilty cases and above 0.5 for the guilty cases for all, experts, and novices. This suggests the participants had a guilty bias in their money laundering assessments compared with the actual trial decisions. In terms of the proportion of correct assessments from Table 2, the novice participants Mean(c) = 0.5122 (where $Mean(c_j) = \frac{1}{jN}\sum_{j=1}^{j} c_{i,j}$) were slightly more successful than the expert participants Mean(c) = 0.5010, but this difference was not statistically significant (p = 0.746) on the basis of the Mann–Whitney U test and the Kruskal–Wallis's test found. The performance accuracy of both expert and novice participants was marginally over 0.5, which indicates almost half of their assessments were incorrect. The guilty bias of the participants could partly have contributed to this average performance because 50% of the cases presented in the vignettes were legitimate transactions not linked to proceeds from crime. The result implies that both novice and expert participants encountered some level of difficulty in distinguishing legitimate transactions from money laundering transactions linked with proceeds from crime.

Fig. 1 shows the graph of the mean probability Mean(r_i) (where $Mean(r_i) = \frac{1}{j}\sum_{i=1}^{j} r_{i,j}$) for the key risk indicator (KRI) analysis for the cases. For Case 1, Case 4, Case 5, Case 6, Case 7, Case 9, Case 10, and Case 11 with key risk indicators designated as high risk for the three (that is customer risk, geographic risk, and transaction risk, where H denoted high and L denotes low in Fig. 1) risk-based assessment criteria, the experts' mean probability was below 70%, whereas such a trend was not reported for the novice participants. Two key reasons make this report particularly relevant. First, it validates the assumption that experts recognise these indicators when making decisions (Simon, 1987). Secondly, it offers evidence that experts' confidence in their judgment decreases as the perceived risk level of a transaction increases. Thus, estimates of key risk criteria as per the statutory AML guideline significantly contributes to AML risks assessment decisions made by experts (the experts sharply distinguished cases with all FATF high risk designated criteria through their perceived lower certainty in their judgment).

Table 2
Mean accuracy measures and comparison of significance *(p)*.

Component measure	Expert	Novice	*p*-value
Overall			
M(c)↑ (proportion correct)	0.5010	0.5122	0.7460
MSPS ↓	0.2967	0.3027	0.9370
Calibration			
Bias (overconfidence) 0	0.1904	0.1896	0.7990
Bias^2(BS)↓	0.0595	0.0567	0.7540
Resolution (discrimination)			
Slope (SL) ↑	0.0159	0.0070	0.6010
Resolution variation (RV) 0	0.2261	0.2328	0.3380
Noise			
Scatter (Error variation) ↓	0.0111	0.0132	0.1870

Pairwise comparisons (expert vs novice) via Mann–Whitney U test and Kruskal–Wallis Test; Definition in text; ↑-larger values better; ↓-smaller values better; 0-zero the best value.

4.2. The Mean Squared Probability Score (MSPS)

In Table 2, the expert participants' MSPS (0.2967) was slightly lower than for the novice participants MSPS (0.3027). In order to test the reliability of group differences by comparing them with those statistics, nonparametric tests were used (Siegel, 1956). Mann–Whitney U test and Kruskal–Wallis's test found no significant difference (*p*-value = .826). Surprisingly the MSPS scores for the experts and novice participants were slightly worse off than the MSPS value of 0.25 for a uniform assessor who assign the same probability to all the possibilities. This is an indication of poor amount of probability assessment skills on the part of the expert and novice participants.

4.3. Specific aspects of accuracy or performance — the statistical decomposition of the MSPS

4.3.1. Calibration

Calibration is one of the most commonly used measures for evaluating the accuracy of judgments expressed in probabilistic form (Keren, 1991). Observing judges' probabilistic assessments, verifying the associated propositions, and then observing the proportion of true responses in each category of response permits us to evaluate judges' calibration empirically (Lichtenstein et al., 1977). The use of a calibration diagram enhances the study of calibration (Yates et al., 1989). In a calibration diagram, all participants' probability estimates are categorised into scores intervals (for example, 0.50–0.59, 0.60–0.69, 0.70–0.79, 0.80–0.89, 0.90–0.99 and 1) and analysed into plots. Probability curves show the relationship between the relative frequency plots of correct answers in each category with the average probability answers for those categories (Harvey, 1997).

Fig. 2 is the calibration diagram for both the expert and the novice participants' judgments. The horizontal axis represents the participants' probability estimates, and the vertical axis defines the relative frequency plots of correct answers in probability estimate category. For example, the top coordinate point for experts (solid lines) in Fig. 2 indicates that 11.7% of their total probability estimates are 100%, and approximately 65% of these estimates were accurate. Novice participants' judgments are plotted in open points (dash lines), while experts' judgments are plotted in filled points (solid lines).

Illustrated in Figs. 3 and 4 are the covariance graphs for the judgments made by the expert and the novice participants. The covariance

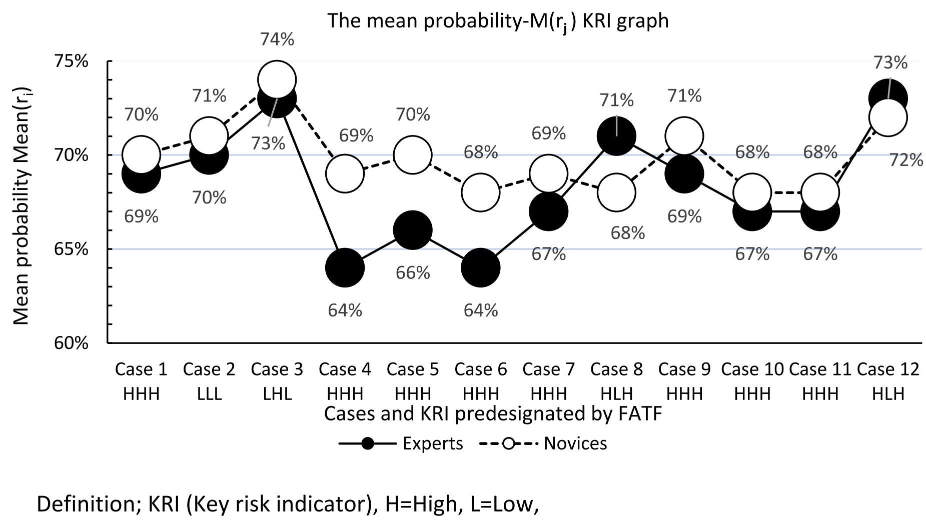

Fig. 1. The mean probability Mean (r_j).

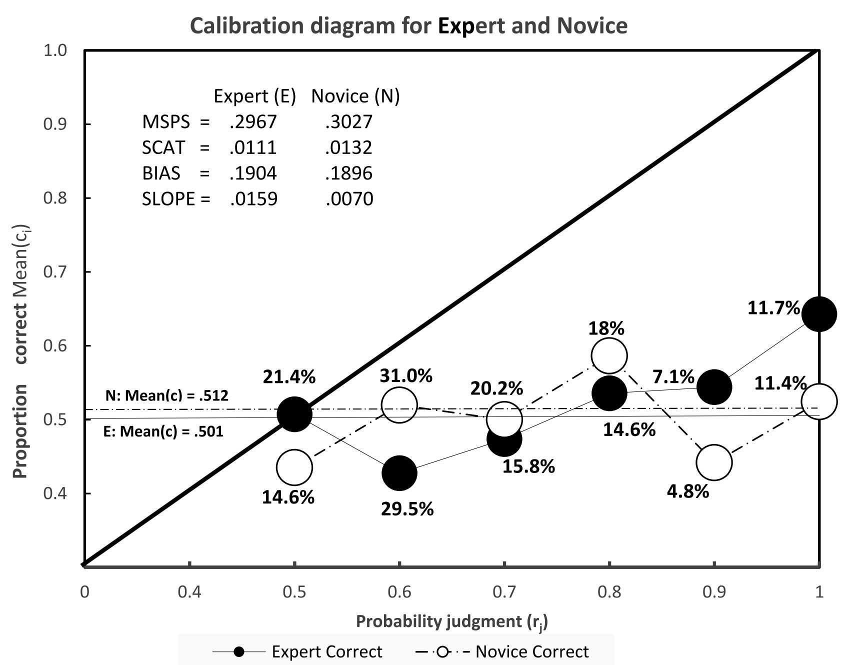

Fig. 2. Calibration diagram for expert and novice.

graph provides an additional virtual illustration approach to probabilistic judgment accuracy component analysis (Yates & Curley, 1985). In Figs. 3 and 4, each covariance graph comprises horizontal and vertical lines along coordinates (0, Mean(r)) and (Mean(c), 0), respectively. Where, $Mean(r) = \frac{1}{JN}\sum_{i=1}^{N}\sum_{j=1}^{J} r_{i,j}$. The horizontal axis is determined by the outcome index. Alternative outcomes resulting from the outcomes index are also identified, i.e., answers that were correctly answered ($c = 1$) or incorrectly answered ($c = 0$). The number in parentheses next to each outcome index value indicates the frequency of occurrence. For instance, in Fig. 3, it is shown that the expert participants selected the correct alternative 481 times but were wrong on 479 instances. Participants' range of probability judgments are described on the vertical axis of each graph.

In Figs. 3 and 4, the distributions shown are proportional histograms. The total sum of the distribution proportions on both hand sides of each graph is 100%. Consider, for example, the histogram on the right-hand side of the expert participants covariance graph. There it is shown that 7.5% of the total 960 judgments made by those experts were probability scores of 100% certainty on the actual correct choice. While on the opposite histogram of the same graph with the incorrect response, 4.2% of the expert participants' total 960 judgments were probability scores of 100% certainty on wrong choices.

Typically, overconfidence is associated with more difficult judgmental tasks, while underconfidence is a form of flawed self-believe that is associated with easier tasks (Lichtenstein et al., 1977). A probabilistic judgment is overconfident to the extent that the bias is positive and large (Yates et al., 1989). In the covariance graph (see Figs. 3 and 4), the intersection point of horizontal and vertical dotted lines is the bias indicator point for each participant group. The horizontal line passes through the mean probability judgment Mean(r). The vertical line goes through the mean outcome index or base rate Mean(c), which is also the proportional correct in the current study. The bias reflects overconfidence (positive) when the intersection is at any point above the 1:1 diagonal, and any point below represents underconfidence (negative) and is perfectly calibrated if it is on the diagonal line (Caycedo-Marulanda et al., 2021).

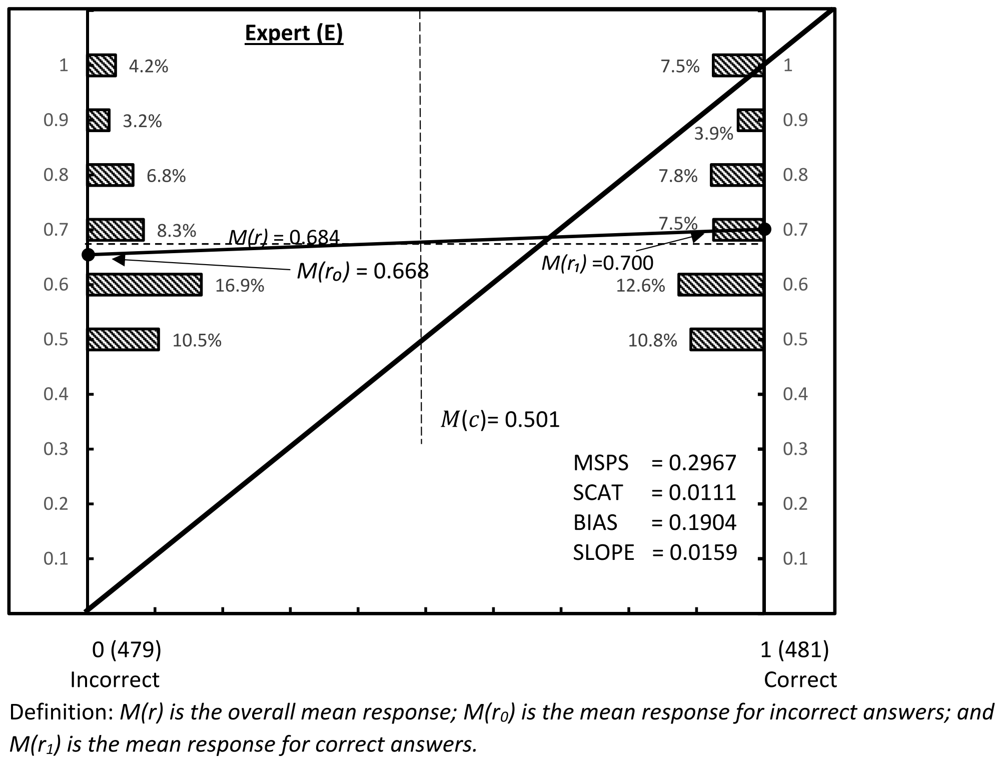

Fig. 3. Expert covariance graph.

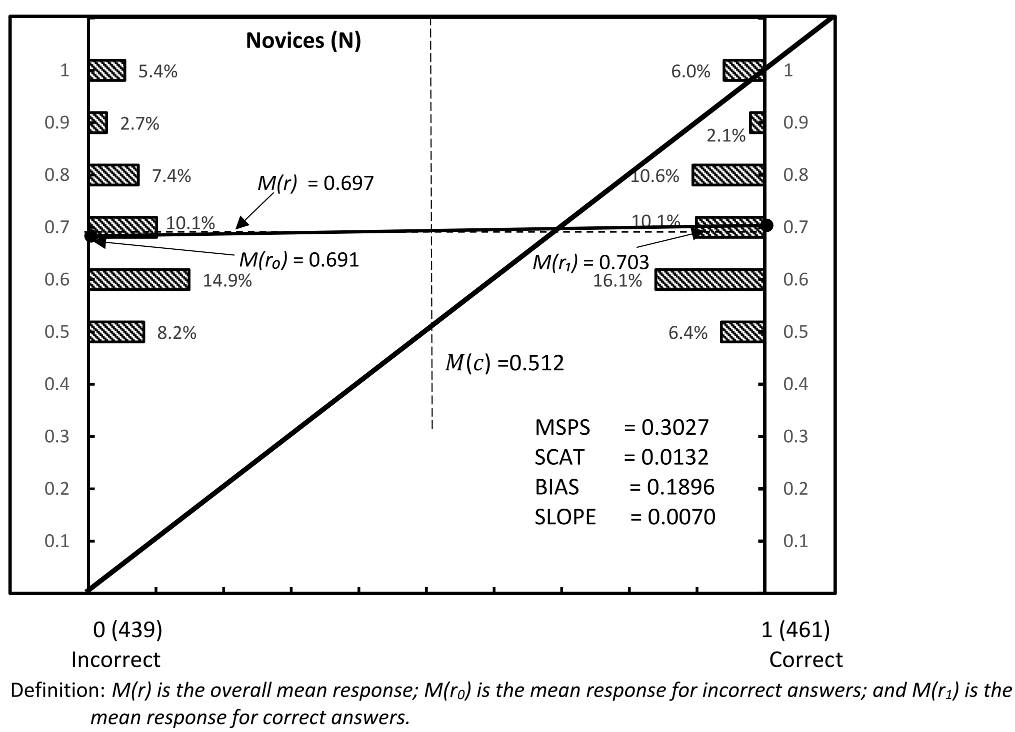

Fig. 4. Novice covariance graph.

Hence, Figs. 3 and 4 show that the intersection point of horizontal and vertical dotted lines for the expert and novice participants falls above the 1:1 diagonal. This result suggests that the expert and novice participants are overconfident in the AML risk assessment (the overconfidence bias). However, the expert (0.1904) participants' bias score appears slightly higher than the novice (0.1896) participants. Thus, indicating that the experts' participants exhibited higher overconfidence in their probability judgment. The comparison of biases for individual participants in the groups was not statistically significant (p-value = 0.799), as indicated in Table 2.

4.3.2. Resolution (Slope)

Basically, the slope index is the regression line for probability judgments regressed on outcomes, passing through the coordinate points

(0, Mean(r_0)) and (1, Mean(r_1)) in the covariance graph (Yates et al., 1989). In Figs. 3 and 4, the virtual inspection of both covariance graphs shows the steeper (better) slope for the expert participants compared to the novice participants. This result suggests that the experts were better able to differentiate, on average, between instances when a financial transaction was likely to result in money laundering crimes and when it was not. However, as shown in Table 2, the distribution comparison of slope values for individual participants in the two groups was not statistically significant (p = 0.601).

An analysis of the calibration diagram also reveals the associated resolution component (RV) via its vertical coordinates of the points (Yates et al., 1989). There is good resolution to the extent that the points are far away from the target event's Mean(c), overall relative frequency. Like the visual trend revealed by the array of points in Fig. 2, the values in Table 1 also indicate RV was higher for novices (0.2328) than for experts (0.2261), thus implying better resolution for experts. Once again, expert participants were more capable of identifying correct assessment decision from incorrect decision. It appears, however, that the RV values for individual participants in the two groups were not statistically significantly different (p-value = 0.3380).

4.3.3. Scatter (error variation)

The final aspect of judgment accuracy discussed in the preliminary framework is scatter (error variation). This measure is based on the noise or scatter in the probabilistic judgments that are not related with the accuracy of the answer (Harvey, 1997). Yates et al. (1989) attributed the causes of error variation to two sources. The first is the judge's inherent inconsistency. Second, it can also occur when a judge is perfectly consistent but relies on cues that are not sufficiently related to the outcome. In Figs. 3 and 4, the dispersion of histograms on either side of the covariance graph represents scatter (error variation). Scatter increases with the degree of dispersion of the distributions. Comparisons between Figs. 3 and 4 show that novice participants had the worst scatter (higher) compared to experts. According to Table 2, there was no statistically significant difference between the scatter values for the two groups of participants (p = 0.187).

5. Discussion and conclusions

5.1. Key findings and implications for research

AML experts are required to make important decisions regarding information that might lead to a suspicion or knowledge of money laundering. Poorly informed decision-making can have a wide range of consequences, from reputational damage to regulatory reprimand and fines (dalla Pellegrina et al., 2022; Gelemerova, 2009). Experts in this domain face the dilemma of maximising sensitivity while reducing false-positives and false-negatives by making reasonable assessments (Maurer, 2005). However, practice variation, i.e., assessing similar activities or transactions differently, remains a recurring issue, although, quality assurance initiatives, such as providing continuing training and publishing guidelines for risk assessment, are taken to address this issue. Assessing money-laundering risk is not an exact risk measurement but embodies the subjective, impressionistic evaluation of the assessor (Riccardi et al., 2019; Sinha, 2014). Given that the decision-making environment is uncertain, expectations do not solely depend on the likelihood of assumptions but also on the expert's level of confidence (Freitas, 2021).

In the AML context, appropriate judgment and confidence levels are vital since under-confidence may lead to denying financial assistance unnecessarily, while overconfidence may lead to trusting and authorising a high-risk offender. Consequently, we investigated expert probability judgments in an AML setting, focusing on calibration and resolution. Considering the expert vs novice comparisons, we draw the following conclusions. First, both experts and novices were biased to label all transactions as suspicious and both groups guilty bias was symmetric. They overwhelmingly preferred false-positive over false-negative errors regardless of transaction perceived likelihood of money laundering risk. In previous research, it was argued that experts do not exhibit the same bias as novices (Bond, 2008; Krems & Zierer, 1994). However, the current results do not support this conclusion for AML risk assessment. Instead, the results are compatible with the conclusion on cognitive biases that expert judgments under uncertainty are susceptible to pretty much the same cognitive biases that novice judgments are susceptible to (Mizrahi, 2013, 2018).

Second, based on proportion accuracy Mean(c) scores, expertise did not significantly affect the capability to distinguish between financial transactions linked or not linked to proceeds of crimes. We found a poor correlation between participant level of expertise and predictive accuracy. Novice participants slightly outperformed expert participants in the proportion of correct answers, despite evidence from the study that expert participants adhered to practice guidelines. Although, Stewart, Roebber, and Bosart (1997) work suggests that apart from personal characteristics, task domain can also affect experts' diagnostic or predictive accuracy. In addition, Phillips, Klein, and Sieck (2004) noted that expert diagnostic or predictive abilities might not be possible in domains with little opportunity for effective feedback, and considering that national FIUs do not regularly provide AML experts with effective feedback regarding suspicious transactions filed (Gelemerova, 2009; Lannoo & Parlour, 2021). This may be compared to weather forecasting, for instance, where accurate and timely feedback are provided regularly or even daily on predictions, providing windows of opportunity for improving certain accuracy dimensions. Consequently, the findings of this study suggest that national FIUs should regularly provide feedback to AML experts regarding the quality of suspicious transaction reports in order to improve their cognitive processes and biases.

Third, novices and experts alike appear to be overconfident about their distribution judgments, and this effect was slightly more pronounced in expert groups. In Lichtenstein et al. (1977) view, overconfidence bias emerges when judgments are made about difficult items. Hence the observed overconfidence bias appears to reflect the reality identified in the literature that it is difficult for experts to distinguish financial transactions that are truly suspicious from those which are not (e.g., Bello & Harvey, 2017). Moreover, experts' judgments about their answers to money laundering suspicious transactions were particularly good at distinguishing between instances when those answers were correct and instances when they were incorrect. Expertise in many settings seems to depend on perception skills, particularly the ability to make good distinctions (e.g., Klein & Hoffman, 2020). Such ability is essentially relevant because they are supposed to be the basis for reporting suspicious transactions. Overall, these findings agree with Mizrahi (2013) conclusion that expert opinions are considerably less accurate than random chances.

5.2. Implications for practice

Our results raise the question of how malefic a bias is if over 85% of experts (89% of novices) are affected? The implications of incorrect risk estimations for financial institution filtering obligations are far-reaching. The overconfidence bias alone affects more than 85% of experts, and if other biases also contribute to flawed risk judgments to a similar extent, then a substantial proportion of AML experts risk assessment decisions may be inaccurate. Moreover, overconfidence is only one cognitive bias among many other biases (e.g., confirmation biases, omission biases, hindsight bias). Given the nature of money laundering and the variability of the underlying criminality, it is difficult to gain a clear picture of the problem, and without which effective action is impossible (Lannoo & Parlour, 2021). In addition, human biases and social consensus seem to be incorporated into the design of AML risk assessments specific to the assumed nature and level of the money laundering risk, as determined by government decision-makers

(Van Duyne et al., 2018). Hence, factors affecting the quality of risk assessment (such as overconfidence) remains an important aspect of scientific interest in this domain.

The experts in our sample preferred false-positive errors over false-negative errors, which may increase inefficiency and expensive costs associated with high false-positive judgments (Amicelle & Iafolla, 2018). But why should the AML experts' judgments have a fairly average proportion accuracy of mean outcome accuracy scores? The ability to discriminate well requires the person reporting judgments to understand how things will turn out. But this is not the case for AML risk assessment domain, where there is little or no feedback from enforcement agencies to AML experts on how well their filed suspicious transactions. Establishing a practice and feedback regimen is one way to facilitate the development of expertise in specific judgments (Phillips et al., 2004). An approach like this is traditional for strengthening skills that can be defined and measurable. However, in the current AML compliance sanction regime, it is possible that even if experts are willing to stick with their intuition on risk judgment, the adverse effects of AML enforcements on noncompliance may skew some risk assessment decision outcomes toward enforcement side. Laws enforcement agencies should bear in mind the difficulties in assessing AML risk when dealing with the regulated entities. If AML compliance officers are under unfair pressure, they will not be working to prevent money laundering, but they will just be protecting themselves (Bello, 2017).

5.3. Limitations and future studies

Our study responds to calls to examine the individual-based role in assessing money laundering risk (Isa et al., 2015) from the perspective of judgment and decision making (Jamil et al., 2022). Our analysis follows Yates (1982), which can be applied when using dichotomous outcome indices, and more generally when using non-dichotomous weighted outcome indices (that requires modifying the outcome index variable, which is not considered here). Yates (1982) undertook a similar decomposition of the MSPS (although he used the term Mean Probability Score, MPS). Yates used alternative formula specifications, but the results are the same in the case of a dichotomous outcome index. Our formula specifications outlined in equations (1 to 6) can also be applied to full-range probability forecasts by substituting $r_{i},_{j}$ for $p_{i},_{j}$ and $c_{i},_{j}$ for e_i in the equations for dichotomous and non-dichotomous empirical probabilities e_i. This could provide an additional method to analyse the results. It can also be useful when analysing composite forecasts or the coherence (consistency) of probability assessments between participants.

Limitations of the study stem from its status as an exploratory examination piece. By attempting to examine the quality of expert judgment, this study illustrates the existence of overconfidence cognitive bias but provides minimal clarification of the mechanisms behind it. Researchers must look beyond the specific effects observed in this study in future studies. For example, the possible cognitive bias by judges in trial decisions. It will also be important to examine the ways in which cultural and gender differences influence risk discrimination and cognitive bias associated with money laundering risk indicators. The experimental design of our study is based on the experimental paradigm of Cantor et al. (2014), that calls for a common module that provides contextual information that is intended to be invariant across a variety of versions of the vignette. To date, its validity has not been proven in the AML risk assessment context, and thus the results of our study should be interpreted cautiously.

Another limitation is the lack of benchmarking of our data and findings with similar studies on risk assessment in other domains. There are studies related to the current research (e.g., those conducted by the C-Rise, Memorial University), and one of these studies estimated the risk of human error in an engineering maintenance context using the success likelihood index method (Noroozi, Khakzad, Khan, MacKinnon, & Abbassi, 2013). Therefore, a direction for future work might be to compare our data and result against those available from that study and other similar research by the C-RISE group on human factors. Additionally, it will be interesting for future studies to explore the impact of feedback on judgment accuracy and compare the performance pre- and post-feedback.

CRediT authorship contribution statement

Henry Ogbeide: Conceptualization, Writing – original draft, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Writing – review & editing. **Mary Elizabeth Thomson:** PhD Project Supervision, Writing – review & editing. **Mustafa Sinan Gonul:** PhD Project Supervision, Conceptualization, Writing – review & editing. **Andrew Castairs Pollock:** Data Curation, Methodology. **Sanjay Bhowmick:** PhD Project Supervision, Implication review. **Abdullahi Usman Bello:** PhD Project Supervision,Writing – review & editing.

Data availability

Data will be made available on request.

Appendix A

See Table A.1.

Appendix B

See Table B.1.

Appendix C

See Table C.1.

Table A.1
Examples of potentially higher and potentially lower money laundering risk factors.
Source: (FATF, 2014).

Factor	Higher risk	Lower risks
Customer risk factor	• The business relationship is conducted in unusual circumstances (e.g., significant unexplained geographic distance between the financial institution and the customer) • Non-resident customers, Legal persons or arrangements that are personal asset-holding vehicles. • Companies that have nominee shareholders or shares in bearer form. • Business that are cash-intensive • The ownership structure of the company appears unusual or excessively complex given the nature of the company's business.	• Financial institutions and DNFBPs — where they are subject to requirements to combat money laundering and terrorist financing consistent with the FATF Recommendations, have effectively implemented those requirements, and are effectively supervised or monitored in accordance with the Recommendations to ensure compliance with those requirements. • Public companies listed on a stock exchange and subject to disclosure requirements (either by stock exchange rules or through law or enforceable means), which impose requirements to ensure adequate transparency of beneficial ownership. • Public administrations or enterprises.

(continued on next page)

Table A.1 (*continued*).

Factor	Higher risk	Lower risks
Country or geographic risk factors	• Countries identified by credible sources, such as mutual evaluation or detailed assessment reports or published follow-up reports, as not having adequate AML/CFT systems. • Countries subject to sanctions, embargos or similar measures issued by, for example, the United Nations. • Countries identified by credible sources as having significant levels of corruption or other criminal activity. • Countries or geographic areas identified by credible sources as providing funding or support for terrorist activities, or that have designated terrorist organisations operating within their country.	• Countries identified by credible sources, such as mutual evaluation or detailed assessment reports, as having effective AML/CFT systems. • Countries identified by credible sources as having a low level of corruption or other criminal activity.
Product, service, transaction or delivery channel risk factors:	• Private banking, Anonymous transactions (which may include cash). • Non-face-to-face business relationships or transactions, Payment received from unknown or un-associated third parties.	• Life insurance policies where the premium is low. • Insurance policies for pension schemes if there is no early surrender option and the policy cannot be used as collateral. • A pension, superannuation or similar scheme that provides retirement benefits to employees, where contributions are made by way of deduction from wages, and the scheme rules do not permit the assignment of a member's interest under the scheme. • Financial products or services that provide appropriately defined and limited services to certain types of customers, so as to increase access for financial inclusion purposes

Table B.1
Summary of vignette used for this study.

Money laundering techniques	Vignette
Bulk cash smuggling	***Case 1.*** Mrs Hussai, a 35-year-old British National, is the sole signatory to PDTransfer Ltd company bank account. This company bank account was opened and used within the UK (United Kingdom) authority. The company runs a service for transmitting money from the UK mainly to Pakistan. The company work by collecting cash from its clients after satisfactory ID verification and then deposit the funds to its bank account before onward payment to the destinated beneficiary in Pakistan. The recent review of PDTransfer Ltd bank records showed that between 1st July 2018 and February 2020, US$3.7 million cash deposits went through the account in more than 400 transactions, and then transferred abroad, principally to a Pakistan based currency exchange business account. PDTransfer Ltd keeps records to show the identities of the various clients from whom the money has been collected in the UK and of those to whom it was ultimately beneficiary in Pakistan.
	Case 2. Mr Evans, a 46-year-old British National, regularly comes into the banking hall to make cheques lodgement into his company account. A first customer due diligence checks conducted by the bank on Evans revealed that he is the proprietor of a motorcycle workshop and garage. And as part of the workshop operating procedure, their customer pays by cheque, which must be lodged in the workshop company bank account. The recent review of Evan's business transactions showed nearly 40 cheques (amount valued between $573 and $6500) deposits totalling over $170,000 within a 3-month period. The record also revealed corresponding debits with transaction narration for; entertainment (amount valued between $50 and $560), dining (amount valued between $70 and $6500), jewellery (amount valued between $573 and $6500), and electronic purchases (amount valued between $106 and $860).
Structuring	***Case 3.*** Miss Abiola, a 25-year-old Nigerian National, regularly comes into the banking hall to make a cash deposit into his account. Due to the way she dresses when she visits the bank, the tellers suspect that she is a peasant (farmworker), but they are not sure of this. On one occasion, a branch teller personally asked Abiola about her occupation, and she became belligerently rude and stopped visiting the branch regularly to make deposits. This change in behaviour prompted one of the bank tellers to review her account. Abiola's bank records indicated her occupation as a student. They also revealed that, prior to her less frequent visits, there was a point (2 months period) where she was depositing in US$ approximately $2200 thrice a month in a variety of lower bills, including $20's, $10's, and $5's.
	Case 4. Dr Bello, a 52-year-old Nigerian National, regularly comes into the banking hall to cash cheques from various customer accounts. A first customer due diligence checks conducted by the bank on Bello revealed that he is the chief medical director of a privately owned hospital. And as part of the hospital operating procedure, their patient billing settlement is subcontracted to a management firm, who in return collects cheque payments from patients and transfer them directly to Dr Bello. The recent review of Bello's transactions records indicated that he usually cashed about twenty to fifty of these cheques cumulatively every ten or eleven days. They also revealed that, within the last 12months period, he cheque-cashed in US$ approximately US$8990 during each visit to the bank.
Virtual currencies	***Case 5.*** Mr Aigbedion, a 35-year-old Nigeria National, is the sole signatory to Tech Ltd company bank account. A first customer due-diligence check on the company profile indicated that the company is a Nigeria-based company that sells encryption services and devices to customers from across the world, and estimated revenue from sales and subscription services exceeded $32 million. The current review of Tech two-year account records revealed that a total of 24 credit inflow of varying value between US$20,000 to US$36,000. The transaction narration notes, "Ongoing subscription fees" and originated from Tech distributors across foreign jurisdiction including the USA, Canada, Australia, Thailand, and the United Arab Emirates. The account balance as at the date of review is US$106,857.57, and the varying sums relating to the 24 debit transactions in the bank account records transferred to 3 different bank accounts owned by crypto exchanges companies.
	Case 6. Mr Adebayo, a 65-year-old Nigerian National, runs a bank account linked to his business interests, which included a convenience store, a property portfolio, and a currency exchange business. Two years ago, the customer-due-diligence report on his business income generation process revealed there are Anti-money laundering system and control in place that fully complied to a satisfactory standard to guide against the inflow of illicit cash into the business. Additionally, Adebayo keeps a proper record of all the business transactions, and a complete history could be ascertained by considering electronic data. However, a recent review of his business account found a US$6,378 transaction relating to a currency exchange deal, and this amount deviates from the average single transaction value of US$2000 occurring in his past transaction history.

(*continued on next page*)

Table B.1 (*continued*).

Misuse of legal entities (Shell companies)	***Case 7.*** Mrs Wards, a 60-year-old British National, completed a one-off debit transfer in US$ approximately US$320,000 from her business account to an offshore jurisdiction (Dubai) account for the purchase of a property with the sum. This company bank account was opened and used within the UK jurisdiction. A first customer-due-diligence check on Wards business activities indicated the source of funding for this transaction came from her trading activities. Her business bank account had no traces of physical-cash deposits, but solely business trade-related transfer payments. Though her business transactions annual turnover exceeds US$1.5 million, this transaction processing officer had concern on the source of funds, because Ward's spouse was a famous businessperson that once held a senior political position in the UK, 10-years ago.
	Case 8. Mrs Hughes, a 37-year-old British National, is the sole signatory to Besco Ltd company bank account. This company bank account was opened and used within the UK jurisdiction. The company runs a diamond trading enterprise. Recent customer due diligence checks on the company profile indicated that Besco Ltd appeared on a national newspaper page, promoting investments with a guaranteed tax-free return of 13.5% per annual. Shortly, after this advertisement, Besco Ltd accounts became active since 2years. And within 3months, the account witness US320,000 credit inflow in 10 transfers from accounts run by Besco Ltd at other local banks domiciled in the UK. The review of Besco Ltd bank records also indicated that Hughes had withdrawn the sum of USDD$171,000 in cash from her company account in twelve debit transactions across the counter within the same period.
Complicit professionals and financial services employees	***Case 9.*** Mrs Adaku, a 35-year-old Nigerian National, is the sole signatory to Coxfx Ltd company bank account. The company runs a service bureau for its clients and has a functioning AML (Anti Money Laundering) unit within its business premises. The company receive cash monies and processed all the cash as payments into and out of the company bank accounts, exchanging US dollar to euro and vice versa. The recent review of Coxfx bank records indicated that between 1st March 2020 and November 2020, US$1.8 million cash deposits went through the accounts in 420 transactions. The ultimate destination of these exchange payments ware paid to three personal accounts owned by the same individual-Adenike Bosede. Coxfx maintains a know your customer file for each client.
	Case 10. Mrs Bosede, a 55-year-old Nigerian National, is the sole signatory to Kunfix Ltd company bank account. This account was opened in 2016 within the UK jurisdiction. The Company runs and trades on the Nigeria money market and has substantial assets. In July 2020, Bosede transferred a sum of about US$3million to Kunfix Ltd bank account. The Fund originated from a bank in Switzerland owned by Bosede. She explained to her UK account manager, that she needed to do this because somebody was trying to gain access to her Switzerland bank account, probably with a view to accessing her account and making unauthorised withdrawals from it. She told the account manager there would only be a short time before she would wish to transfer the sum back to the Switzerland bank account. In late August 2020, Bosede indicated that she wanted to return the money to her Switzerland bank account and enquired when she would be able to do so.
Trade-based money laundering	***Case 11.*** Mr Martins, a 39-year-old British National, is the sole signatory to a personal bank account opened on 6 March 2020. On 6 March and 4 April 2020, Martins deposited on each occasion the sum of US$90,000. A first customer-due-diligence check on Martin indicated the source of funding for these deposits came from sales of properties. The recent review of his bank records (due to the statutory policy on continuous on-going customer-due-diligence) revealed an outflow transfer of $95,000 from his account on 30 April 2020 to another bank account domiciled in a foreign jurisdiction (United Arab Emirates) owned by an individual. This fund transferred originated from the first two consecutive cash deposits of $90,000, and the account balance is $85,000 (30 September 2020) as at the date of this review.
	Case 12. Mr. Davis, a 65-year- old British National, recently bought a luxury car worth US$55,500. He funded the purchase partly through a five-year loan of US$40,000 from a UK commercial bank and paid the balance US$15,500 in cash. A first customer-due diligence check on his source of income, indicated his occupation as the sole owner of a car dealership showroom, and the motor company predicted annual turnover is US$1 million. Further credit checks revealed that Davis had utilised similar loans schemes within the last five years, for six luxury cars procurements. Davis opted for early repayments of these loans in cash within six months of loan disbursement.

Table C.1
Summary of vignette key risk indicators.

Money laundering techniques	Vignettes	Key risk criteria	AML/ money laundering indicators	Outcome
Bulk cash smuggling	Case 1	Customer Risk-High Geographic Risk-High Transaction Risk - High	PDTransfer Ltd keeps records to show the identities of the various clients from whom the money has been collected in the UK and of those to whom it was ultimately beneficiary in Pakistan.	Non-conviction
	Case 2	Customer Risk-Low Geographic Risk-Low Transaction Risk - Low	The record also revealed corresponding debits with transaction narration for; entertainment (amount valued between $50 and $560), dining (amount valued between $70 and $6500), jewellery (amount valued between $573 and $6500), and electronic purchases (amount valued between $106 and $860).	Convicted
Structuring	Case 3	Customer Risk-Low Geographic Risk-High Transaction Risk - Low	Abiola's bank records indicated her occupation as a student. Her record also revealed that, there was a point (2 months period) where she was depositing in US$ approximately $2200 thrice a month in a variety of lower bills, including $20's, $10's, and $5's.	Non-conviction
	Case 4	Customer Risk-Low Geographic Risk-High Transaction Risk - Low	The recent review of Bello's transactions records indicated that he usually cashed about twenty to fifty of these cheques cumulatively every ten or eleven days. They also revealed that, within the last 12months period, he cheque-cashed in US$ approximately US$8990 during each visit to the bank.	Convicted

(*continued on next page*)

Table C.1 (*continued*).

Money laundering techniques	Vignettes	Key risk criteria	AML/ money laundering indicators	Outcome
Virtual currencies	Case 5	Customer Risk-High Geographic Risk-High Transaction Risk - High	Varying sums relating to the 24 debit transactions in the bank account records transferred to 3 different bank accounts owned by crypto exchanges companies.	Convicted
	Case 6	Customer Risk-High Geographic Risk-High Transaction Risk - High	The customer-due-diligence report on his business income generation process revealed there are Anti-money laundering system and control in place that fully complied to a satisfactory standard to guide against the inflow of illicit cash into the business. Additionally, Adebayo keeps a proper record of all the business transactions, and a complete history could be ascertained by considering electronic data.	Non-conviction
Misuse of legal entities (Shell companies)	Case 7	Customer Risk-High Geographic Risk-Low Transaction Risk - Low	A first customer-due-diligence check on Wards business activities indicated the source of funding for this transaction came from her trading activities. Her business bank account had no traces of physical-cash deposits, but solely business trade-related transfer payments.	Non-conviction
	Case 8	Customer Risk-High Geographic Risk-Low Transaction Risk - High	The company runs a diamond trading enterprise. Recent customer due diligence checks on the company profile indicated that Besco Ltd appeared on a national newspaper page, promoting investments with a guaranteed tax-free return of 13.5% per annual. Shortly, after this advertisement, Besco Ltd accounts became active since 2years.	Convicted
Complicit professionals and financial services employees	Case 9	Customer Risk-High Geographic Risk-High Transaction Risk - High	The recent review of Coxfx bank records indicated that between 1st March 2020 and November 2020, US$1.8 million cash deposits went through the accounts in 420 transactions. The ultimate destination of these exchange payments ware paid to three personal accounts owned by the same individual-Adenike Bosede. Coxfx maintains a know your customer file for each client.	Convicted
	Case 10	Customer Risk-Low Geographic Risk-Low Transaction Risk - Low	The Company runs and trades on the Nigeria money market and has substantial assets.	Non-conviction
Trade-based money laundering	Case 11	Customer Risk-High Geographic Risk-High Transaction Risk - High	A first customer-due-diligence check on Martin indicated the source of funding for these deposits came from sales of properties.	Non-conviction
	Case 12	Customer Risk-High Geographic Risk-Low Transaction Risk - High	Further credit checks revealed that Davis had utilised similar loans schemes within the last five years, for six luxury cars procurements. Davis opted for early repayments of these loans in cash within six months of loan disbursement.	Convicted

Table D.1
Vignettes participants country of origin.

Participants	Country of residence	Frequency	Percent
AML professional	Nigeria	48	60.0
	UK	18	22.5
	USA	4	5.0
	India	1	1.3
	Singapore	1	1.3
	Cyprus	1	1.3
	Italy	1	1.3
	Malta	1	1.3
	Portugal	1	1.3
	UAE	2	2.5
	Russia	1	1.3
	Ghana	1	1.3
	Total	80	100.0
Novice	Nigeria	50	66.7
	UK	18	24.0
	India	6	8.0
	Pakistan	1	1.3
	Total	75	100.0

Appendix D

See Table D.1.

References

Afflerbach, P., van Dun, C., Gimpel, H., Parak, D., & Seyfried, J. (2021). A simulation-based approach to understanding the wisdom of crowds phenomenon in aggregating expert judgment. *Business & Information Systems Engineering*, *63*(4), 329–348.

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, *17*(4), 351–371.

Alexander, C. S., & Becker, H. J. (1978). The use of vignettes in survey research. *Public Opinion Quarterly*, *42*(1), 93–104.

Amicelle, A., & Iafolla, V. (2018). Suspicion-in-the-making: Surveillance and denunciation in financial policing. *The British Journal of Criminology*, *58*(4), 845–863.

Anichebe, U. (2020). Combating money laundering in an age of technology and innovation. Available at SSRN 3627681.

Atzmuller, C., & Steiner, P. M. (2010). Experimental vignette studies in survey research.. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *6*(3), 128.

Axelrod, R. M. (2017). Criminality and suspicious activity reports. *Journal of Financial Crime*, *24*(3), 461–471.

Baldwin, G. (2003). The new face of money laundering. *Journal of Investment Compliance*.

Bell, R. E. (2000). Proving the criminal origin of property in money-laundering prosecutions. *Journal of Money Laundering Control*, *4*(1), 12–25.

Bello, A. U. (2017). *Improving anti-money laundering compliance: Self-protecting theory and money laundering reporting officers*. Springer.

Bello, A. U., & Harvey, J. (2017). From a risk-based to an uncertainty-based approach to anti-money laundering compliance. *Security Journal*, *30*(1), 24–38.

Benjamin, D., Mandel, D. R., & Kimmelman, J. (2017). Can cancer researchers accurately judge whether preclinical reports will reproduce? *PLoS Biology*, *15*(6), Article e2002212.

Berg, A. (2020). The identity, fungibility and anonymity of money. *Economic Papers: A Journal of Applied Economics and Policy*, *39*(2), 104–117.

Bergström, M., Svedberg Helgesson, K., & Mörth, U. (2011). A new role for for-profit actors? The case of anti-money laundering and risk management. *JCMS: Journal of Common Market Studies*, *49*(5), 1043–1064.

Bond, G. D. (2008). Deception detection expertise. *Law and Human Behavior*, *32*(4), 339–351.

Borracci, R. A., & Arribalzaga, E. B. (2018). The incidence of overconfidence and underconfidence effects in medical student examinations. *Journal of Surgical Education*, *75*(5), 1223–1229.

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3.

Canhoto, A. I. (2021). Leveraging machine learning in the global fight against money laundering and terrorism financing: An affordances perspective. *Journal of Business Research*, *131*, 441–452.

Cantor, D. E., Blackhurst, J. V., & Cortes, J. D. (2014). The clock is ticking: The role of uncertainty, regulatory focus, and level of risk on supply chain disruption decision making behavior. *Transportation Research Part E Logistics and Transportation Review*, *72*, 159–172.

Carter, K., Sabers, D., Cushing, K., Pinnegar, S., & Berliner, D. C. (1987). Processing and using information about students: A study of expert, novice, and postulant teachers. *Teaching and Teacher Education*, *3*(2), 147–157.

Caulkins, J. P., & Reuter, P. (2022). How much demand for money laundering services does drug selling create? Identifying the key parameters. *International Journal of Drug Policy*, *103*, Article 103652.

Caycedo-Marulanda, A., Patel, S. V., Verschoor, C. P., Chadi, S. A., Moslein, G., Raval, M., et al. (2021). Comparing 'Twitter'polls results with an online survey on surgeons perspectives for the treatment of rectal cancer. *BMJ Innovations*, *7*(1).

Cindori, S., et al. (2013). Money laundering: correlation between risk assessment and suspicious transactions. *Financial Theory & Practice*, *37*(2), 181–206.

Costello, F., & Watts, P. (2014). Surprisingly rational: probability theory plus noise explains biases in judgment.. *Psychological Review*, *121*(3), 463.

dalla Pellegrina, L., Di Maio, G., Masciandaro, D., & Saraceno, M. (2022). Are bankers "crying wolf"? Type I, type II errors and deterrence in anti-money laundering: The Italian case. *Italian Economic Journal*, 1–29.

Demetis, D. S. (2010). *Technology and anti-money laundering: A systems theory and risk-based approach*. Edward Elgar Publishing.

Demetis, D. S., & Angell, I. O. (2007). The risk-based approach to AML: representation, paradox, and the 3rd directive. *Journal of Money Laundering Control*.

Endress, A. D. (2013). Bayesian learning and the psychology of rule induction. *Cognition*, *127*(2), 159–176.

FATF (2014). *Risk-based approach guidance for the banking sector*: *Techreport*, Financial Action Task Force.

FATF (2022). *Data pooling, collaborative analytics, and data protection-phase 2*: *Techreport*, Paris, France: Financial Action Task Force.

Fedirko, T. (2021). Suspicion and expertise: following the money in an offshore investigation. *Journal of the Royal Anthropological Institute*, *27*(1), 70–89.

Freitas, G. V. R. (2021). Narrative economics and behavioral economics: contributions to the behavioral insights on post-Keynesian theory. *Brazilian Journal of Political Economy*, *41*, 372–384.

Garcia, J., Gomez, Y., & Vila, J. (2022). Financial overconfidence, promotion of financial advice, and aging. *Journal of Business Research*, *145*, 325–333.

Gelemerova, L. (2009). On the frontline against money-laundering: the regulatory minefield. *Crime, Law and Social Change*, *52*(1), 33–55.

Halliday, T., Levi, M., & Reuter, P. (2019). Anti-money laundering: an inquiry into a disciplinary transnational legal order. *UC Irvine J. Int'L Transnat'L & Comp. L.*, *4*, 1.

Harvey, N. (1997). Confidence in judgment. *Trends in Cognitive Sciences*, *1*(2), 78–82.

Isa, Y. M., Sanusi, Z. M., Haniff, M. N., & Barnes, P. A. (2015). Money laundering risk: from the bankers' and regulators perspectives. *Procedia Economics and Finance*, *28*, 7–13.

Jamil, A. H., Mohd-Sanusi, Z., Mat-Isa, Y., & Yaacob, N. M. (2022). Money laundering risk judgement by compliance officers at financial institutions in Malaysia: the effects of customer risk determinants and regulatory enforcement. *Journal of Money Laundering Control*.

Johari, R. J., Zul, N. B., Talib, N., & Hussin, S. A. H. S. (2020). Money laundering: Customer due diligence in the era of cryptocurrencies. In *1st International conference on accounting, management and entrepreneurship* (pp. 130–135). Atlantis Press.

Kabir, S., & Papadopoulos, Y. (2019). Applications of Bayesian networks and Petri nets in safety, reliability, and risk assessments: A review. *Safety science*, *115*, 154–175.

Keren, G. (1991). Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychologica*, *77*(3), 217–273.

Kerlinger, F. N. (1966). *Foundations of behavioral research*. New York.

Klein, G. A., & Hoffman, R. R. (2020). Seeing the invisible: Perceptual-cognitive aspects of expertise. In *Cognitive Science Foundations of Instruction* (pp. 203–226). Routledge.

Krems, J. F., & Zierer, C. (1994). Are experts immune to cognitive bias? Dependence of" confirmation bias" on specialist knowledge. *Zeitschrift Fur Experimentelle Und Angewandte Psychologie*, *41*(1), 98–115.

Kruisbergen, E. W., Leukfeldt, E. R., Kleemans, E. R., & Roks, R. A. (2019). Money talks money laundering choices of organized crime offenders in a digital age. *Journal of Crime and Justice*, *42*(5), 569–581.

Lannoo, K., & Parlour, R. (2021). *Anti-money laundering in the EU: Time to get serious*. Centre for European Policy Studies.

Larson, J. S., & Billeter, D. M. (2017). Adaptation and fallibility in experts' judgments of novice performers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(2), 271.

Levi, M., & Soudijn, M. (2020). Understanding the laundering of organized crime money. *Crime and Justice*, *49*(1), 579–631.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1977). Calibration of probabilities: The state of the art. *Decision Making and Change in Human Affairs*, 275–324.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980," judgment under uncertainty: heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*.

Linn, C. J. (2010). Redefining the bank secrecy act: Currency reporting and the crime of structuring. *Santa Clara Law Review*, *50*, 407.

Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition? *Psychological Science*, *24*(12), 2351–2360.

Maurer, B. (2005). Due diligence and "reasonable man," offshore. *Cultural Anthropology*, *20*(4), 474–505.

McKenzie, C. R., Liersch, M. J., & Yaniv, I. (2008). Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Processes*, *107*(2), 179–191.

McLaughlin, J. S., & Pavelka, D. (2013). The use of customer due diligence to combat money laundering. *Accountancy Business and Public Interest*, 57–84.

Mizrahi, M. (2013). Why arguments from expert opinion are weak arguments. *Informal Logic*, *33*(1), 57–79.

Mizrahi, M. (2018). Arguments from expert opinion and persistent bias. *Argumentation*, *32*(2), 175–195.

Muradoglu, G., & Onkal, D. (1994). An exploratory analysis of portfolio managers' probabilistic forecasts of stock prices. *Journal of Forecasting*, *13*(7), 565–578.

Murphy, A. H. (1972a). Scalar and vector partitions of the probability score: Part I. Two-state situation. *Journal of Applied Meteorology (1962-1982)*, 273–282.

Murphy, A. H. (1972b). Scalar and vector partitions of the probability score: Part II. N-state situation. *Journal of Applied Meteorology and Climatology*, *11*(8), 1183–1192.

Murray, J., & Thomson, M. E. (2010). Applying decision making theory to clinical judgements in violence risk assessment. *Europe's Journal of psychology*, *6*(2), 150–171.

Musharraf, M., Hassan, J., Khan, F., Veitch, B., MacKinnon, S., & Imtiaz, S. (2013). Human reliability assessment during offshore emergency conditions. *Safety science*, *59*, 19–27.

Naheem, M. A. (2017). Trade based money laundering: exploring the implications for international banks.

Naheem, M. A. (2019). Anti-money laundering/trade-based money laundering risk assessment strategies–action or re-action focused? *Journal of Money Laundering Control*.

Noroozi, A., Khakzad, N., Khan, F., MacKinnon, S., & Abbassi, R. (2013). The role of human error in risk analysis: Application to pre-and post-maintenance procedures of process facilities. *Reliability Engineering & System Safety*, *119*, 251–258.

Nyreröd, T., Andreadakis, S., & Spagnolo, G. (2022). Money laundering and sanctions enforcement: large rewards, leniency and witness protection for whistleblowers. *Journal of Money Laundering Control*.

Olsson, H. (2014). Measuring overconfidence: Methodological problems and statistical artifacts. *Journal of Business Research*, *67*(8), 1766–1770.

Paddam, A., Barnes, D., & Langdon, D. (2010). Constructing vignettes to investigate anger in multiple sclerosis. *Nurse Researcher*, *17*(2).

Parr, W. V., Heatherbell, D., & White, K. G. (2002). Demystifying wine expertise: Olfactory threshold, perceptual skill and semantic memory in expert and novice wine judges. *Chemical Senses*, *27*(8), 747–755.

Phillips, J. K., Klein, G., & Sieck, W. R. (2004). Expertise in judgment and decision making: A case for training intuitive decision skills. *Blackwell Handbook of Judgment and Decision Making*, *297*, 315.

Pol, R. F. (2020). Anti-money laundering: The world's least effective policy experiment? Together, we can fix it. *Policy Design and Practice*, *3*(1), 73–94.

Poulou, M. (2001). The role of vignettes in the research of emotional and behavioural difficulties. *Emotional and Behavioural Difficulties*, *6*(1), 50–62.

Raza, M. S., Zhan, Q., & Rubab, S. (2020). Role of money mules in money laundering and financial crimes a discussion through case studies. *Journal of Financial Crime*.

Reyna, V. F., & Lloyd, F. J. (2006). Physician decision making and cardiac risk: effects of knowledge, risk perception, risk tolerance, and fuzzy processing.. *Journal of Experimental Psychology: Applied*, *12*(3), 179.

Riccardi, M., Milani, R., & Camerini, D. (2019). Assessing money laundering risk across regions. An application in Italy. *European Journal on Criminal Policy and Research*, *25*(1), 21–43.

Sanders, N. R., & Ritzman, L. P. (1992). The need for contextual and technical knowledge in judgmental forecasting. *Journal of Behavioral Decision Making*, *5*(1), 39–52.

Savona, E. U., & Riccardi, M. (2017). *Assessing the risk of money laundering in Europe. Final report of project IARM*. Transcrime–Joint Research Centre on Transnational Crime.

Schneider, S. (2020). *Money laundering in British Columbia: A review of the literature*. The Cullen Commission of Inquiry into Money Laundering in British Columbia.

Siegel, S. (1956). Measures of correlation and their tests of significance. New York: McGraw-Hill,

Simon, H. A. (1987). Making management decisions: The role of intuition and emotion. *Academy of Management Perspectives*, *1*(1), 57–64.

Singh, K., & Best, P. (2019). Anti-money laundering: using data visualization to identify suspicious activity. *International Journal of Accounting Information Systems*, *34*, Article 100418.

Sinha, G. (2014). To suspect or not to suspect: Analysing the pressure on banks to be 'Policemen'. *Journal of Banking Regulation*, *15*(1), 75–86.

Stewart, T. R., Roebber, P. J., & Bosart, L. F. (1997). The importance of the task in analyzing expert judgment. *Organizational Behavior and Human Decision Processes*, *69*(3), 205–219.

St.Marie, B., Jimmerson, A., Perkhounkova, Y., & Herr, K. (2021). Developing and establishing content validity of vignettes for health care education and research. *Western Journal of Nursing Research*, *43*(7), 677–685.

Takáts, E. (2011). A theory of "crying wolf": The economics of money laundering enforcement. *The Journal of Law, Economics, & Organization*, *27*(1), 32–78.

Trueblood, J. S., Holmes, W. R., Seegmiller, A. C., Douds, J., Compton, M., Szentirmai, E., et al. (2018). The impact of speed and bias on the cognitive processes of experts and novices in medical image decision-making. *Cognitive Research: Principles and Implications*, *3*(1), 1–14.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185*(4157), 1124–1131.

Van Duyne, P. C., Harvey, J. H., & Gelemerova, L. Y. (2018). *The critical handbook of money laundering: Policy, analysis and myths*. Springer.

Veen, H., Heuts, L., & Leertouwer, E. (2020). *Dutch national risk assessment on money laundering 2019*. WODC.

Wallsten, T. S., & Budescu, D. V. (1983). State of the art—Encoding subjective probabilities: A psychological and psychometric review. *Management Science*, *29*(2), 151–173.

Wilkes, C. J. (2020). A case for reforming the anti-money laundering regulatory regime: How financial institutions' criminal reporting duties have created an unfunded private police force. *Ind. LJ*, *95*, 649.

Yates, J. F. (1982). External correspondence: Decompositions of the mean probability score. *Organizational Behavior and Human Performance*, *30*(1), 132–156.

Yates, J. F., & Curley, S. P. (1985). Conditional distribution analyses of probabilistic forecasts. *Journal of Forecasting*, *4*(1), 61–73.

Yates, J. F., McDaniel, L. S., & Brown, E. S. (1991). Probabilistic forecasts of stock prices and earnings: The hazards of nascent expertise. *Organizational Behavior and Human Decision Processes*, *49*(1), 60–79.

Yates, J. F., Zhu, Y., Ronis, D. L., Wang, D.-F., Shinotsuka, H., & Toda, M. (1989). Probability judgment accuracy: China, Japan, and the United States. *Organizational Behavior and Human Decision Processes*, *43*(2), 145–171.

Zavoli, I., & King, C. (2021). The challenges of implementing anti-money laundering regulation: An empirical analysis. *The Modern Law Review*, *84*(4), 740–771.

Henry Ogbeide is a Ph.D. student at Newcastle Business School, Northumbria University. His research focuses on integrating statistics and psychology in the formation and evaluation of decision-making. A stream of his research explores the quality of anti-money laundering risk assessments, such as the quality of professional probability judgment.

Mary E. Thomson is Professor of Decision Science at Newcastle Business School. Her research focuses on financial judgmental forecasting, expert judgment and decision making in a variety of contexts, risk perception and risk assessment. She completed her Ph.D. in 1999 and has since published in a wide variety of book and journals, including the European Journal of Operational Research, the International Journal of Forecasting, Risk Analysis, Decision Support Systems, Journal of Behavioural Decision Making, Journal of Empirical Legal Studies, the International Journal of Forensic Mental Health, Legal and Criminological Psychology, Journal of Customer Behaviour, and the European Management Journal.

M. Sinan Gönül is an Associate Professor of Business Analytics and Acting Deputy Head of the Department at the Newcastle Business School, Northumbria University. He completed his Ph.D. in Business Administration (in the area of Decision Science & Operations Management) in 2007. Prior to his Ph.D., he holds an MBA and a B.Sc.(Eng) in Electrical and Electronics Engineering. He also conducted postdoctoral research on judgment and decision making in the Department of Psychology at Glasgow Caledonian University between 2009–2010. His research focuses on judgment & decision-making, judgmental forecasting, decision and forecasting support systems and behavioural operational research/operations management. He has published in various prestigious journals including European Journal of Operational Research, Decision Sciences Journal, Journal of Behavioural Decision Making, Decision Support Systems, Production and Operations Management, International Journal of Forecasting, Journal of Forecasting and Technological Forecasting & Social Change.

Andrew C. Pollock is an independent statistical analyst who was formally a Reader at Glasgow Caledonian University and retired in 2008. He completed his Ph.D. in 1988 and since then has published numerous articles in a variety of books and journals. His research interests focus on integrating economics, finance, statistics and psychology in the formation and evaluation of point and probability forecasts, particularly price movements in currency, equity and economic series using judgemental, time series and econometric techniques.

Sanjay Bhowmick is Associate Professor and Associate Head of Business at Northeastern University London. His research interests lie in firm internationalisation, entrepreneurship, social cause venturing, management of the commons, and strategy. He has published in and reviewed for international journals and has been a member of international research bodies including as a core committee member of the Australian Centre of Entrepreneurship Research Exchange annual conference series and a founding member of Academy of Management Entrepreneurship Division International Committee.

Dr. Abdullahi U. Bello (FCCA, CFE, CAMS) is a fellow of the ACCA, a Certified Fraud Examiner (CFE) and a is a Fellow of Higher Education UK. He has over 20 years' experience in banking, law enforcement, consulting and the academia. He holds a Ph.D. in Anti-Money Laundering (Northumbria), an M.Sc. in Forensic Accounting (Sheffield Hallam) and a B.Sc. in Applied Accounting (Oxford Brookes). He also holds a BTech in Computer Science and a Postgraduate in Higher Education Practice. Dr Bello is an Assistant Professor in the Accounting and Financial Management Department of Newcastle Business School, Northumbria University. He is the Programme Leader of M.Sc. Forensic Accounting at the University.