Received: 21 September 2017 | Revised: 26 June 2018 | Accepted: 27 June 2018

DOI: 10.1111/ijau.12137

SPECIAL ISSUE ARTICLE

WILEY

Do fine feathers make a fine bird? The influence of attractiveness on fraud-risk judgments by internal auditors

Marc Eulerich[1] | Jochen C. Theis[1] | Junpeng Lao[2] | Meike Ramon[2]

[1] Mercator School of Management, University of Duisburg–Essen, Duisburg, Germany

[2] Department of Psychology, University of Fribourg, Fribourg, Switzerland

Correspondence
Marc Eulerich, University of Duisburg–Essen, Mercator School of Management, Chair for Internal Auditing, Lotharstraße 65, 47057 Duisburg, Germany.
Email: marc.eulerich@uni-due.de

Independence and objectivity are key principles assumed to underlie internal auditors' fraud-risk judgments. However, a substantial body of evidence suggests that physical attractiveness of suspects may influence internal auditors' fraud-risk judgments. In this experimental study we investigated whether internal auditors are susceptible to appearance-related biases, or whether they correct for them due to their expertise and motivation. A total of 193 internal auditors were presented a misappropriation-of-assets scenario in which the attractiveness of a suspect was manipulated. To determine whether professional expertise is associated with increased resilience to appearance-related biases, their fraud-risk judgments were contrasted with those acquired from 240 subjects without auditing experience ("naive subjects"). In line with our predictions, attractiveness modulated naive subjects' fraud-risk judgments, whereas internal auditors did not show any indication for appearance-related biases. Our findings suggest that internal auditors' experience and motivation may immunize them to the phenomena of physical attractiveness stereotyping and the attractiveness halo effect.

KEYWORDS

audit failure, ethics, fraud, internal audit, professionalism

1 | INTRODUCTION

A perpetual problem for the management of all types of companies is the misappropriation of assets and fraud. According to a study of the Association of Certified Fraud Examiners (ACFE), organizations lose 5% of their annual revenue due to asset misappropriation, corruption, and financial statement fraud, with more than one-fifth of the cases causing losses of at least $1 million for each respective company (ACFE, 2012); for comparable statistics, see also Beasley, Carcello, Hermanson, and Neal (2010) and Beasley, Carcello, Hermanson, and Lapides (2000). Consequently, fraud detection and prevention have recently become significant concerns, especially for the top management and internal auditors (IAs)—top management's support unit (Abbott, Park, & Parker, 2000; Cooper, Dacin, & Palmer, 2013; Kaplan, Pope, & Samuels, 2010; Moeller, 2004; Norman, Rose, & Rose, 2010; Prawitt, Sharp, & Wood, 2012). Typically, internal auditing acts as a direct agent of the chief executive officer (CEO) and the audit committee. One can define internal auditing as follows (Institute of Internal Auditors [IIA], 2013):

> *Internal auditing is an independent, objective assurance and consulting activity designed to add value and improve an organization's operations. It helps an organization accomplish its objectives by bringing a systematic, disciplined approach to evaluate and improve the effectiveness of risk management, control, and governance processes.*

There is a growing literature showing the value of internal auditing for management, especially in fraud-risk judgments[1] and detection (ACFE, 2012; Norman et al., 2010).

According to the IIA Standards, the paramount principles of IAs' work are independence and objectivity (IIA, 2013). Hence, in a potential fraud case an IA should be objective in his or her fraud-risk judgments (AICPA, 2002; IIA/AICPA, 2008) and should not consider any

characteristic of a suspect, such as physical attractiveness or other individual traits (for details see IIA, 2013, standard 1120; see also Mutchler, 2003).[2]

However, assessing the probability that an individual will commit or has committed fraud or a similar person-related decision often involves (direct or indirect) human interaction between the IA and the individual being assessed (for related research, also see Fanning & Piercey, 2014). Importantly, various studies suggest a relationship between judgments of a person and their physical appearance or attractiveness (Alley, 1988; Berscheid, 1986; Jackson, 1992). Therefore, the physical appearance of an auditee could also directly affect an IA's ability to operate in an independent and objective manner and consequently poses a potential threat to the value of internal auditing for management.

Thus, in the present study we aimed to empirically determine the impact of physical attractiveness on IAs' fraud-risk judgments. Our goal was to assess whether IAs operate objectively and independently, and therefore in accordance with the two main pillars of the whole internal audit profession. We specifically sought to determine whether both principles are jeopardized by appearance-related biases established previously across a range of social scenarios. To address this outstanding question, we opted for an interdisciplinary approach; specifically, we integrated insights and state-of-the-art methodology from corporate governance and auditing, as well as psychology and visual cognition. Inter alia, we make recourse to and extend existing literature on the question as to whether potential defendants benefit from physical attractiveness.

Integrating knowledge and tools from these areas, we designed a quasi-lab between-subjects experiment conducted with large samples of IAs and naive subjects with no auditing experience. In this setting, participants received information about a hypothetical company with indicators of decreasing profits and the potential for fraud to have had occurred (misappropriation-of-assets scenario).[3] This case information was accompanied by a stylized personnel file of the employee who is potentially responsible for the suspected fraud, including a photograph of the employee's face. Importantly, the depicted faces were computer-generated stimuli created for this unique purpose, based on independent viewers' attractiveness ratings of real faces, and involving state-of-the-art computational approaches as used in vision research (see Section 3). Unbeknownst to participants, the face stimuli included in the personnel file varied in attractiveness, age, and gender.

Participant groups consisted of highly experienced IAs, as well as naive subjects lacking comparable professional expertise, and therefore ability and motivation to correct for judgment biases. To anticipate our findings, in line with the physical attractiveness stereotype and attractiveness halo effect (see Section 2), we observed that naive subjects made less favorable fraud-risk judgments for *unattractive* suspects. In contrast, as expected based on the implications of the process model of appearance-based stereotyping, our findings suggest that IAs corrected for potential judgment biases as their fraud-risk judgments seemed unaffected by suspects' attractiveness.

Our results support the ability of IAs to make objective judgments, one of their main characteristics, and give new insights for the understanding of the profession. Since most of the prior internal audit research focuses on the impairment of an IA's objectivity, our empirical results present a positive view of the profession. Using a multidisciplinary approach with state-of-the-art techniques from psychology, we open a small part of the behavioral black box of auditor judgments. Our findings emphiasize the high quality of experienced IAs compared to naive subjects, which can be accredited to the value of IAs training and professional experience.

2 | THEORETICAL BACKGROUND

2.1 | Physical attractiveness, responsibility, and IA judgment

Socialization, social expectancy, and evolutionary perspectives imply that "more physically attractive individuals either inherently possess or come to develop more positive personality traits" (Lorenzo, Biesanz, & Human, 2010, p. 1777; see also Little, 2014, & Langlois et al., 2000). This consistent tendency to attribute more positive qualities (e.g., social skills, potency, adjustment and intellectual ability, and trustworthiness; Adams, 1982; Dion, Berscheid, & Walster, 1972; Feingold, 1992; Langlois et al., 2000; Miller, 1970a; Rohner & Rasmussen, 2011) to people who are physically attractive rather than unattractive constitutes the *physical attractiveness stereotype* and the *attractiveness halo effect* (Joseph, 1982). Both seem to operate from an early age onwards (Dion, 1973; Dion & Berscheid, 1974; Gross & Crofton, 1977), and suggest that a core element of human socialization is learning the stereotype "what is beautiful is good" (Dion et al., 1972), or, in other words, "fine feathers make a fine bird" (Patzer, 2006). Most definitions of physical attractiveness or appearance refer to facial appearance as opposed to body appearance. This is reasonable, because "the face is usually the first […] and often the most potent source of information […] during social interaction" (Jackson, 1992, p. 3). Definitions of facial appearance are manifold and, for example, include factors such as age and gender (Alley, 1988, p. 2). However, a core element seems to be attractiveness (Bashour, 2007, p. 16), the evaluation of which is characterized by "remarkable consensus" (Jackson, 1992, p. 4); that is, high inter-rater consistency. Consequently, in our study we chose to manipulate facial attractiveness as argued in Section 3.

Miller (1970b) argued that physically attractive individuals are perceived as having an internal locus of control and hence are considered "as individuals who are not easily influenced or manipulated by others, and whose opinions spring from independent thinking and personal conviction, all of these being qualities which denote greater […] credibility," or, in other words, trustworthiness (for further reading, also see Joseph, 1982). Trustworthiness is of specific interest for our study because it affects, monitors, and guides individuals' actions and attitudes in their interactions and is a subjective state (Kasper-Fuehrera & Ashkanasy, 2001). Specifically, perceived trustworthiness reduces suspicion and increases openness toward an individual (Shinners, 2009; Szulanski, Cappetta, & Jensen, 2004). Gomulya, Wong, Ormiston, and Boeker (2017) analyzed the effects of facial appearance on CEO selection and argue that the facial appearance is a signal for trustworthiness. Also, similar to gender typicality biases reported for hiring decisions (Von Stockhausen, Koeser, & Sczesny,

2013), physical attractiveness has been found to bias managers' reactions to applications and résumés in favor of attractive candidates (Marlowe, Schneider, & Nelson, 1996; Przygodzki-Lionet, Olivier, & Desrumaux, 2010; Quereshi & Kay, 1986). Furthermore, attractive communicators advocating a specific opinion across a range of settings (e.g., attitude toward speed limits on highways) have been demonstrated to induce comparatively larger changes in opinion than unattractive communicators (Chaiken, 1979; Horai, Naccari, & Fatoullah, 1974; Rule & Tskhay, 2014; Snyder & Rothbart, 1971).

Sigall and Ostrove (1975) reported that attractive defendants received lower ratings of guilt and less severe recommendations for punishment than unattractive defendants in a simulated jury scenario (Gross & Crofton, 1977; for comparable results see also Efran, 1974; Michelini & Snodgrass, 1980). Similarly, Leventhal and Krate (1977) found a significantly negative relationship between a defendant's physical attractiveness and assigned length of sentencing, and Mazzella and Feingold (1994) argued that it is generally advantageous for defendants to be physically attractive. To summarize, the outlined findings suggest that attractive suspects would be less likely judged as guilty during an audit.

2.2 | Judgment bias and correction for IAs

Considering this substantial body of empirical evidence, one question remains: If human information processing in general is threatened by physical attractiveness, are there reasons to believe that IAs are immune to it? A substantial line of empirical evidence suggests that biases in judgments following the physical attractiveness stereotype might be corrected to different degrees, depending on the assessors' cognitive capacities (low vs. high ability) and their belief that it is inappropriate to rely on physical appearance (low vs. high motivation) (Bodenhausen & Macrae, 1998; Hart, Ottati, & Krumdick, 2011; Rahn & Cramer, 1996; Wegener & Petty, 1997). For instance, Hart et al. (2011) found that voters generally evaluated attractive political candidates more favorably than otherwise equivalent unattractive candidates (also see Efran & Patterson, 1974). However, others have reported that this effect can be reduced (undercorrected), accurately corrected, or even overcorrected (reversed) provided participants dispose of high ability and motivation (also see Bodenhausen & Macrae, 1998; Rahn & Cramer, 1996; Wegener & Petty, 1997). Correspondingly, Marlowe et al. (1996) reported that hiring decisions were least biased in favor of attractive candidates for the most experienced managers in their sample.

Based on such findings, Hart et al. (2011) proposed a simple process model of appearance-based stereotyping (for similar conceptualizations, also see Bodenhausen & Macrae, 1998; Wegener & Petty, 1997). According to this model, an individual's physical appearance is initially *categorized* (attractive vs. unattractive), followed by *activation* of the physical attractiveness stereotype. This stereotype is *applied* when forming other judgments of this person. Finally, the assessor "corrects for the perceived influence of the physical attractiveness stereotype" (Hart et al., 2011). In cases for which the magnitude of the initial assimilation effect is underestimated/correctly estimated/overestimated, the assimilation effect will be reduced/eliminated/reversed. In the latter case (reversal), "an unattractive [person] will be rated more positively than an (otherwise equivalent) attractive [person]" (Hart et al., 2011). However, the correction of an assimilation effect requires an adequate amount of cognitive resources (high ability), because correction only occurs if the "perceiver identifies and becomes aware of the biasing influences of appearance" (Hart et al., 2011). Furthermore, the perceiver needs to believe that it is inappropriate to rely on physical appearance in judgments (high motivation).

While IAs may not differ from people with similar social/economic status or perspectives in terms of general disposition or inherent intelligence, they usually possess expert knowledge, growing with work experience in internal auditing. In addition, professional standards require continuous professional training. Both aspects are relevant to IAs' judgments in general and, given the specific setting of our experiment, fraud-risk judgments in particular (for related research, see Zimbelman, 1997). According to the aforementioned model, the average IA in our experiment should therefore possess *sufficient ability* to correct for biases. Furthermore, IAs' *motivation* to do so should be very high, given the paramount principle of objectivity conveyed in the field's professional standards and training.[4] Consequently, we assume that IAs' attempt to correct for the perceived influence of the attractiveness halo effect on their judgments in general and, given the specific setting of our experiment, fraud-risk judgments in particular.[5] Taking into consideration previous findings (e.g., Bodenhausen & Macrae, 1998; Rahn & Cramer, 1996; Wegener & Petty, 1997), we acknowledge that their judgments might also be affected in the *opposite* manner; that is, their extensive experience and motivation for objectivity could actually lead to an *overcorrection* for attractiveness-related biases. That is, physical attractiveness may influence, and therefore jeopardize, objectivity of (fraud-risk) judgments when professional experience and motivation are lacking, as well as when abundant. Consequently, based on the outlined theoretical framework, we differentiate between and assess fraud-risk judgments in a situation in which professional experience and motivation are lacking (naive subjects) and when abundant (IAs).

For naive subjects with no auditing experience, we formulate:

> **H1.** Higher physical attractiveness of suspects will be negatively related to attributed fraud responsibility.

For IAs, we formulate:

> **H2.** IAs will correct their judgments, which are biased by physical attractiveness, to fulfill their objectivity requirements.

3 | METHODS

3.1 | Stimulus creation

We employed a state-of-the-art objective procedure for the creation of face stimuli varying in perceived attractiveness. This procedure entailed three steps described in the following subsections: (a) *initial attractiveness rating* for young and old face stimuli taken from a larger database (for details, see later and Table 1); (b) *creation* of unattractive/attractive younger and older female/male faces based on (a); (c) *validation* of perceived attractiveness of face stimuli created in (b) by an independent

TABLE 1 Demographic data on two sets of faces rated for attractiveness

	Younger faces (*n* = 84)			Older faces (*n* = 69)		
	n	Mean age ± *SD*	Age range	n	Mean age ± *SD*	Age range
Female	46	26.4 ± 6.4	20–40	38	63.2 ± 9.5	44–86
Male	38	26.4 ± 6.6	20–40	31	58.3 ± 11.2	41–79

group of observers. Thus, our procedure identified and validated the facial information determining attractiveness ratings across a large group of observers and faces (see also Oosterhof & Todorov, 2008; Torrance, Wincenciak, Hahn, DeBruine, & Jones, 2014), in a data-driven manner.

This three-step procedure was adopted for two main reasons. On the one hand, it is extremely parsimonious; compared with the alternative of identifying face images unequivocally considered attractive, it requires fewer ratings/observers (e.g., Fruhen, Watkins, & Jones, 2015). Second, the stimulus generation procedure is objective: The facial information identified as determining attractiveness ratings was subsequently used to create maximally unattractive/attractive face stimuli (i.e., artificial faces) in a data-driven fashion (b). The final validation step (c) involved obtaining attractiveness ratings from another independent sample to ensure that the stimuli created in (b) are appropriate for the research instrument described later (i.e., the unattractive/attractive stimuli created are actually perceived as such).

3.1.1 | Rating procedure

Four groups of stimuli (old/young, female/male) were selected from a larger, continuously expanding database of 3D face models developed at the Institute of Neuroscience and School of Psychology at the University of Glasgow, UK (see Table 1 for demographic details of the stimulus groups selected for the initial rating procedure). Individual face models in this database are created by capturing the likeness of volunteer subjects using multiple cameras to recreate 3D face models using shape and texture information, which can then be selectively manipulated (for a demonstration, see https://youtu.be/Va-SXrn0_9Q). Stimuli from this database (for examples, see Figure 1) have been and continue to be used to address a range of different research objectives. These include, for example, the development of frameworks for perceptually valid automatic facial expression generation (Yu, Garrod, Jack, & Schyns, 2015), the role of culture and motion with regard to perception of facial expressions (Jack, Garrod, Yu, Caldara, & Schyns, 2012; Richoz, Jack, Garrod, Schyns, & Caldara, 2015; Yu, Garrod, & Schyns, 2012), as well as the effect of race on face processing proficiency (Ramon et al., 2016).

Four different sets of face stimuli (see Table 1 for details and Figure 1 for examples) were rated for perceived attractiveness. Each set was rated by an independent group of observers (students from a German university; see Table 2 for raters' demographic data). The rating procedure was implemented as an online survey to maximize the number of ratings per stimulus set, with students pseudorandomly assigned to a set (to obtain comparable numbers of female/male raters for each stimulus group). The basic procedure for this attractiveness rating involved central presentation of a randomly selected image from the set of to-be-rated full-frontal 2D images (rendered from original 3D face models using 3D Studio Max; number of trials equals *n* of stimuli per set). Raters were required to provide a subjective attractiveness rating using a seven-point Likert scale (1: very unattractive; 7: very attractive). As indicated earlier, male and female, and younger and older faces were tested separately (four separate rating tasks; see Table 2); each rater was presented one of these subcategories.

3.1.2 | Creation of unattractive/attractive faces for the internal auditing experiment

After obtaining the attractiveness rating for each of the four stimulus samples, a regression technique was used to generate attractive and unattractive male and female faces per age group. Each vertex of the 3D model and each pixel of the texture map were first subtracted from the mean vertex position and pixel value over the set of identities, and these deviation values were then linearly regressed against the attractiveness ratings over the identities per gender and age. Attractive and unattractive face models were finally generated for each gender and age by evaluating the regression equations at extreme points of the input scale and adding the output vertex and pixel deviation values to the average vertex and pixel values. Using

FIGURE 1 Examples of original face stimuli rated for attractiveness. The figure displays examples of a female (left) and male (right) young face. Both individuals displayed here provided consent to publish an image of their likeness [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Demographic data of samples of raters per stimulus group

		Stimulus group rated			
		Young females	Young males	Old females	Old males
Number of raters	Male	14	13	14	13
	Female	18	17	17	20
	Total	32	30	31	33
Raters' age (mean ± SD[age range])	Male	25 ± 2 [22–29]	26 ± 3 [20–30]	26 ± 4 [20–33]	29 ± 5 [22–43]
	Female	26 ± 3 [21–33]	25 ± 3 [21–33]	27 ± 5 [18–38]	26 ± 4 [19–33]
	All	26 ± 3 [21–33]	25 ± 3 [20–33]	27 ± 4 [18–38]	27 ± 5 [19–43]

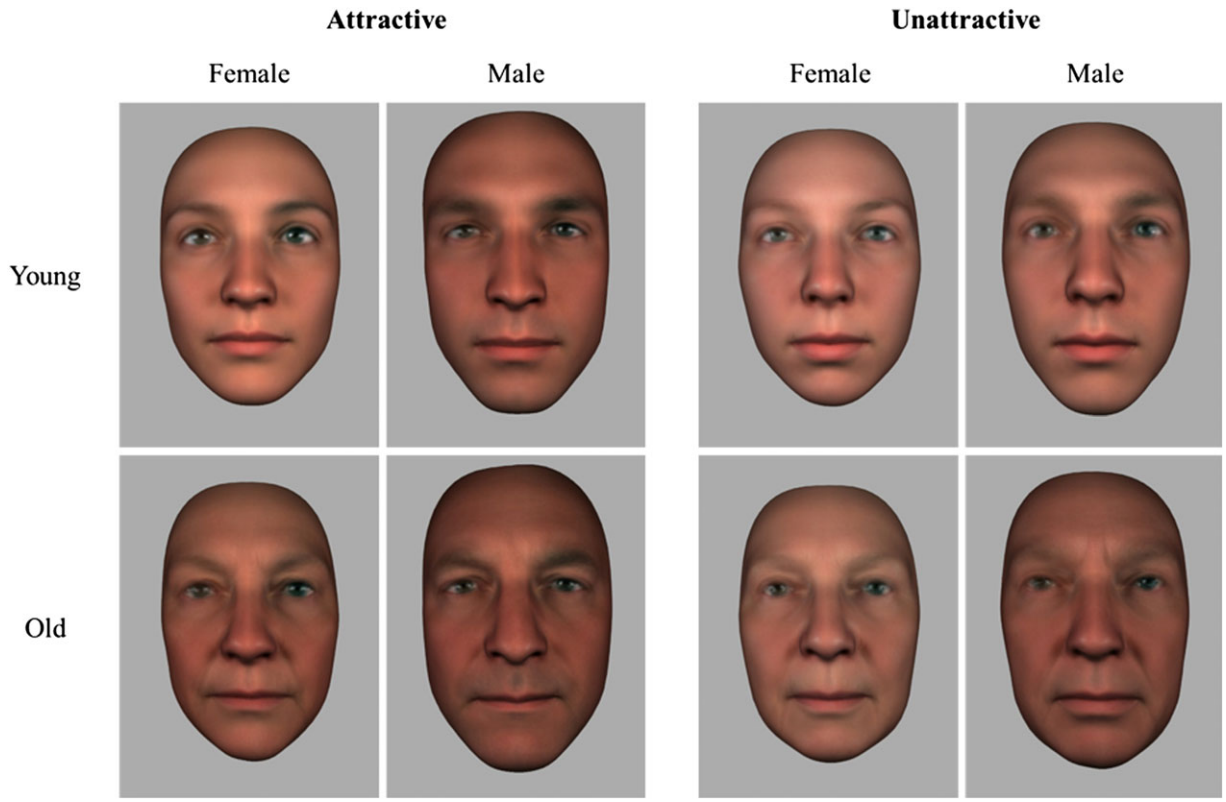


FIGURE 2 Final stimuli (differing in attractiveness, age, and gender) created for the experiment. Note that the stimuli displayed here were generated based on the attractiveness ratings obtained for the original stimuli derived from a larger database (see Section 3 for details and Figure 1 for examples) [Colour figure can be viewed at wileyonlinelibrary.com]

this procedure, we created the final faces used for the experiment (described below), which are displayed in Figure 2. Again, these final stimuli do not depict "actual" individuals, but represent faces that are generated in an objective manner using the facial information that correlated with observers' ratings of attractiveness.

3.1.3 | Validation of unattractive/attractive faces generated for the internal auditing experiment

Finally, we sought to verify that our *attractiveness* manipulation (as implemented via the steps described in Sections 3.1.1 and 3.1.2) produced stimuli that were indeed perceived as unattractive/attractive. To this end, the stimuli generated in Section 3.1.2 were rated by independent observers from Amazon Mechanical Turk (n = 80, half female; average age: 34.14 years, range: 18–63 years). Each observer was presented with one stimulus pair at a time (four pairs in total, random order of presentation). Each pair consisted of direct counterparts (e.g., attractive vs. unattractive) of a given gender. Participants were asked to indicate which of the two faces is more attractive. The observed response pattern suggests that the participants attributed differences in attractiveness to our stimuli in line with our intentions. Specifically, the stimuli that were generated to depict attractive faces were categorized as attractive significantly more than those generated to depict unattractive faces (old males: 70/80; young females: 65/80; old females: 63/80; young males: 56/80).

3.2 | Participants of the experiments

3.2.1 | IAs

The IAs who participated occupied different professional positions and varied in terms of years of experience in and outside the internal audit function (IAF) and qualification. These participants were randomly selected at the annual meeting of the German Chapter of the IIA and were randomly presented with one version of the experimental materials (see Section 3.3 for details). The attendees of the meeting thereby constitute a (relatively) random sample of the population of IAs in Germany. Because the conference qualifies as professional

training, IAs' main motivation to participate is to fulfill their annual professional training requirements.[6]

Of the 825 participants at the conference, 178 IAs were recruited for the experiment; an additional 15 responses were gathered during a training session for IAs, resulting in an initial sample of 193 subjects. Demographic information for the initial sample is provided in Table 3. Sixty IAs were in possession of a specific professional certificate or qualification, with Certified Internal Auditor being the most frequently named. IAs held various positions within the IAF, ranging from junior auditors to chief audit executives (CAEs). The final sample of IAs considered for statistical analyses comprised 182 participants, as only data from IAs who completed the entire experiment (i.e., gave responses to all the questions) were included.

3.2.2 | Naive subjects

This group consists of 240 Bachelor's and Master's degree students from a German university (mean age of 24.64 years, 52.50% females; see Table 3). As students in Germany generally enroll in a Master's program immediately after completion of their Bachelor's degree, our sample comprises individuals with marginal or no professional experience. All students participated in courses related to accounting, auditing, and corporate governance. The experimental materials (Section 3.3) were randomly distributed among students before or after class and were identical to those completed by the IAs (only the demographic questions differed). The final sample of naive subjects considered for statistical analyses comprised data from 214 students (26 observations eliminated due to incomplete responses).

3.3 | Experimental materials

We developed a comprehensive research instrument involving a marginally modified teaching case on internal auditing from H. Snyder, Clifton, and Bowlin (2012) to address our research question. This case (see Appendix) illustrates a situation in which a possible fraud case might have affected the profitability of a fictitious company. The case provides participants with a transcribed conversation between the company's CEO and the CAE. Throughout the dialogue, the CEO provides the CAE information concerning the company's economic situation and other key (operating) figures that indicate possible fraud in the production. The case was translated from English to German (and vice versa to check for possible translation errors); where necessary, the actors' names were replaced with names more common in Germany, and the currency of all amounts was changed from US dollars to euros. In a short instruction section at the beginning of the research instrument, participants are instructed to carefully read the dialogue between the CEO and the CAE in combination with the remaining experimental materials and to assume that he or she is the company's CAE.

To capture the expected influence of a suspect's physical attractiveness on a participant's fraud-risk judgment, the CEO indicates in the dialogue that the production manager *Schmidt* is in charge of the production and that *Schmidt* might be a good starting point for further investigation. However, because we intended to leave some leeway for the participants' own assessments, the CEO's assertions are somewhat vague. To capture the construct physical attractiveness, we manipulated the independent variable attractiveness (attractive vs. unattractive) of the employee who is potentially responsible for the suspected fraud (*Schmidt*) by providing participants with a stylized personnel file of *Schmidt*. Although attractiveness clearly is of most interest to our study, as it constitutes the core element of the theory we apply, *Schmidt's* gender (female vs. male) and age (young vs. old) was also varied to determine whether these factors interact. Consequently, participants were presented with one of eight (2 × 2 × 2) different stimuli/employee's faces each (between-subjects design). Because our interest and argumentation focus on attractiveness, and because neither gender nor age turn out to interact with attractiveness in the analyses, we chose to collapse the analyses as well as the presentation of results and compare the two groups of attractive versus unattractive stimuli/employee's faces in the following. The personnel file contained (among other information) a large (5.6 in. × 4.2 in.) high-definition color photograph of the employee's face (for the creation of the eight stimuli/employee's faces applied in the experiment, see previous sections).

3.4 | Experimental procedure

After reading the case described earlier and viewing the personnel file, participants answered a number of questions. Instructed to assume that *they are the company's CAE*, they were required to answer several questions from this perspective. The first three questions asked addressed different aspects of the participants' judgment and decision-making problem. Participants were asked to (1) assess whether the risk associated with the issue at hand was sufficiently high that further fraud investigation was necessary, (2) make a judgment about the risk level of the audit case, and (3) assess the probability that the production manager *Schmidt* is responsible for the potential fraud. The first two questions captured participants' general risk assessments based on the objective facts provided in the case, whereas we focus on the third question related directly to the suspect and which therefore represents our dependent variable submitted to statistical analyses.

To control for differences in disposition, we also asked the participants to indicate the importance of IAs making objective judgments.

TABLE 3 Demographic data of participants

	Internal auditors				Naive subjects	
	n [%]	Work years mean ± *SD*	Work years as IA mean ± *SD*	Age mean ± *SD*	*n* [%]	Age mean ± *SD*
Male	137 [71]	19 ± 9	11 ± 8	44 ± 9	114 [47]	25 ± 3
Female	36 [19]	15 ± 12	10 ± 10	40 ± 11	126 [53]	24 ± 2
No information	20 [10]	—	—	—	—	—
Overall	193	18 ± 10	11 ± 8	43 ± 10	240	25

As participants unequivocally indicated that objectivity is of very high importance, this variable was omitted from further analyses. Finally, in addition to providing demographic details, participants were asked to answer a number of questions as potential manipulation checks. Apart from the first question (see earlier) for which a binary yes/no answer was required, participants provided answers using a seven-point Likert scale (from −3 to +3) with different endpoints (see Table 4 for general risk assessments and dependent variable and see the Appendix for the remaining questions).

The experimental materials were arranged in a separate folder to ensure that the excerpt from the personnel file was constantly visible to the participants during their work on the case. Each participant was orally instructed that it was crucial for the study that (1) participants carefully follow all instructions in the experimental materials, (2) participants read and answer the questions in the questionnaire one by one, and (3) given answers are not altered after reading subsequent questions. We did not implement stronger controls (e.g., splitting up the set of questions outlined earlier to different envelopes which have to be opened and sealed in a specific sequence), because it was essential to keep overall time investment and complexity of instructions as low as possible. Based on prior experience with other experiments and our pretests, we concluded that stronger controls could significantly decrease the number of participants willing and able to complete the experiment (see earlier for details).

TABLE 4 Dependent variable and general risk assessments and respective questions

Question	Scale endpoint
General risk assessments	
Necessity:	
Is the risk related to the issue at hand so high that further fraud investigation is necessary?	[yes; no]
Risk:	
How high is the risk related to the issue at hand?	[very low; very high]
Dependent variable	
Responsibility:	
How do you assess the probability that Mr./Ms. Schmidt is responsible for the issue/possible fraud?	[very improbable; very probable]

We argue that the research design described constitutes the optimal solution in weighing different important aspects. First, we believe that objectivity in stimulus creation takes priority over the alternative of attempting to select real faces that are considered unattractive/attractive in the population (see Section 3.1). An objective approach also stipulates that the pictures of *Schmidt* show faces without hair (and ears), because haircut and hair color in particular can interact with the perception of attractiveness. Consequently, we did not seek recourse to photographs of real people for our final stimuli but rather utilized face stimuli generated in an *objective*, *data-driven* manner for our specific purpose. Second, while our research idea necessarily involves the presentation of a picture, we had to decide about the salience of the stimulus. We attempted to present the picture of *Schmidt* in a perceptible but natural way, and hence to provide the participants with an extract from *Schmidt's* personnel file. We acknowledge that the participants might suspect that the picture presented relates to the study's intention. However, most importantly, there is no reason to assume that the participants anticipate the study's aim (assessment of the attractiveness stereotype) or expected results, as crucially each participant was only confronted with a single face stimulus.

3.5 | Analyses

Statistical analyses were performed in the programming language and software environment R (R Core Team, 2016) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) and results were displayed using ggplot2 (Wickham, 2009). To test our hypotheses, we fitted a linear mixed model with *Schmidt's* manipulated physical appearance (i.e., *attractiveness*; attractive or unattractive) and *Group* (IAs or naive subjects) as independent variables to predict the subjective responses regarding *Schmidt's responsibility*. Moreover, we used the *Group* predictor as a random intercept to further account for nested random effect within each group.

TABLE 5 Descriptive statistics and pairwise comparison with *t*-tests for both experimental groups

	General risk assessments			Dependent variable	
	Necessity				
Experimental group (observations)	Yes	No	Risk Mean (*SD*)	Responsibility Mean (*SD*)	Rated attractiveness Mean (*SD*)
Internal auditors					
Attractive (93)	59	34	1.06 (1.33)	0.27 (1.26)**	−0.01 (0.95)***
Unattractive (89)	49	40	0.85 (1.17)	−0.03 (1.22)**	−0.43 (1.03)***
Overall (182)	108	74	0.96 (1.25)	0.12 (1.25)	−0.21 (1.01)
Naive subjects					
Attractive (106)	93	13	1.08 (1.17)	0.03 (1.25)***	−0.49 (1.37)**
Unattractive (108)	92	16	1.24 (0.90)	0.57 (1.05)***	−0.74 (1.24)**
Overall (214)	185	29	1.16 (1.04)	0.30 (1.18)	−0.62 (1.31)*

Means differ significantly between groups (one-tailed *t*-test) at:

*10% significance level. **5% significance level. ***1% significance level.

The table is based on observations where there are no data missing.

TABLE 6 Linear mixed model fitting output

Formula: *Responsibility ~ FaceAttr × GROUP + (1| GROUP)*			
Restricted maximum likelihood criterion at convergence: 1,271.1			
Random effects			
Groups	**Name**	**Variance**	**SD**
GROUP	(Intercept)	0.007698	0.08774
Residual		1.430204	1.19591
Number of observations: 396; groups: 2			
Fixed effects			
	Estimate	**SE**	***t*-value**
(Intercept)	0.2688	0.1519	1.770
Attractiveness_Unattr	−0.3025	0.1773	−1.706
GROUP_naive_subjects	−0.2405	0.2104	−1.143
Attractiveness_Unattr: GROUP_naive_sub	0.8390	0.2412	3.478
Correlation of fixed effects			
	(Intercept)	**Attractiveness_Unattr**	**GROUP_naive_sub**
Attractiveness_Unattr	−0.571		
GROUP_naïve_subjects	−0.722	0.412	
Attractiveness_Unattr: GROUP_naive_subjects	0.420	−0.735	−0.569

Table based on observations where there are no data missing.

4 | RESULTS

Table 5 provides descriptive statistics for the responses provided by IAs and naive subjects regarding the dependent variable and general risk assessments in the conditions reflecting our manipulation.[7] Because neither gender nor age of the potential suspect turn out to interact with attractiveness in the analyses, we chose to collapse the analyses as well as the presentation of results and compare the two groups of attractive vs unattractive stimuli/employee's faces.

The linear mixed model was fitted using restricted maximum likelihood estimation. Information of the model fitting, including the coefficient of the fixed effect and the additional variance caused by the random effect, is reported in Table 6. We then performed hypotheses testing on the coefficients of the linear mixed model. An *F*-test on the contrast of the coefficients revealed a strong interaction between *Schmidt's* attractiveness and *Group*, F(1, 392) = 12.10, $p = 0.00056$, whereas both the main effects of *Schmidt's attractiveness* and *Group* were not significant ($F = 1.61$ and $F = 1.03$ respectively).[8]

The interaction and simple effects were further estimated using bootstrap confidence intervals (CIs). As shown in Figure 3, IAs provided mean *responsibility* ratings for both attractive and unattractive stimuli that did not significantly differ from zero (attractive: 0.27, CI = −0.02, 0.56; unattractive: −0.03, CI = −0.33, 0.25). Naive subjects, however, provided *responsibility* ratings that differed as a function of *attractiveness*: Average ratings for attractive stimuli were also around zero (0.03, CI = −0.24, 0.29), whereas unattractive stimuli were considered to be more responsible (0.57, CI = 0.33, 0.79). The magnitude of the increased attribution of *responsibility* was on average a half point on the seven-point Likert scale. Summarizing, regarding our dependent variable *responsibility*, our results are in line with H1 and H2. Naive subjects attributed greater certainty of guilt to physically unattractive suspects, and vice versa attributed lower certainty of guilt to physically attractive suspects. On the other hand, we do not find evidence that IAs had the tendency to judge suspects depending on their attractiveness.

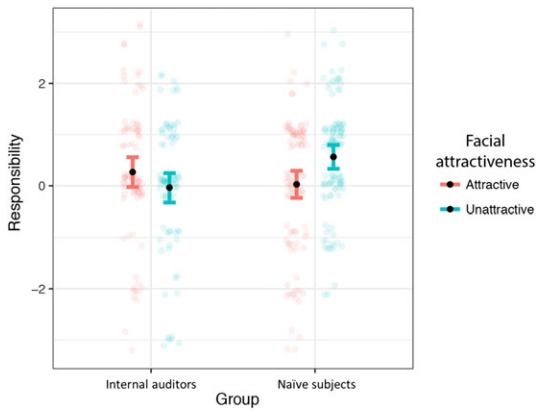


FIGURE 3 Visualization of results obtained in the misappropriation-of-assets scenario. The raw data of each individual participant are shown as transparent dots using a scatter plot. Black solid dots show the estimation of the mean, with the 95% bootstrap confidence intervals shown as error bars. Red color indicates the attractive condition, whereas the light blue color indicates the unattractive condition [Colour figure can be viewed at wileyonlinelibrary.com]

5 | DISCUSSION AND CONCLUSION

Responding to an increased public concern about the level of fraud within organizations, fraud detection and fraud prevention have become increasingly important for internal auditing and corporate governance. In this study we sought to empirically address the question of whether physical attractiveness of a suspect influences IAs'

fraud-risk judgments. From an IIA perspective, whose standards aim for objective (fraud-risk) judgments, we argue that IAs should not be biased by attractiveness (attractiveness halo effect or any other appearance-related characteristics) of a suspect in an audit or fraud case. However, the physical attractiveness stereotype suggests that attractiveness of a suspect may indeed affect the judgment of IAs. This would call into doubt IAs' objectivities and eventually their ability to support the management with objective judgments about an audit object. However, since the IIA Standards and the IIA Code of Ethics promote the (individual) objectivity as one of the main pillars of an IA, training and professional certifications strongly focus on objectivity aspects (e.g., the Certified Internal Auditor exam). Since an IAF cannot be objective, when not every single IA is objective on his or her own, guaranteeing the highest level of individual objectivity is a main task for the chief audit executive. Providing individual and collective professional qualification is a well-recognized way to create a common ground about (individual) objectivity and the tasks of an auditor. Together with experience gained through prior audits, IAs might thus be able to correct for the perceived influence of the attractiveness halo effect on their judgments.

In our misappropriation-of-assets scenario, we manipulated the independent variable *attractiveness of a suspect*. Responses of IAs with varying degrees of experience were compared with those of naive subjects. The latter were assumed to exhibit similar inherent intelligence and social/economic perspectives but—given their lack of professional training and experience—significantly less ability and motivation to correct for biases in judgments compared with IAs.

Regarding our dependent variable *responsibility*, our results are in line with H1 and H2. Naive subjects attributed greater certainty of guilt to physically unattractive suspects and, vice versa, attributed lower certainty of guilt to physically attractive suspects. On the other hand, we do not find evidence that IAs had the tendency to judge suspects depending on their attractiveness. The latter outcome can be explained by the process model of appearance-based stereotyping. In particular, IAs seem to have sufficiently high cognitive capabilities (high ability, expertise) and—given the IIA framework and its principle of objectivity—high motivation to correct for the perceived influence of the attractiveness halo effect when making fraud-risk judgments. In summary, our results indicate that physical attractiveness influences, and therefore jeopardizes, objectivity of fraud-risk assessments given insufficient professional experience (naive subjects). On the other hand, this hazard might be nullified provided there is sufficient motivation and expertise, as indicated by the data obtained for the IA participants.

We give new insights for the IA profession, especially since most prior research solely focuses on identifying factors that impair an IA's objectivity, and in this context, for example, support the attractiveness halo effect. Our findings differ from this view. With our multidisciplinary approach and the techniques utilized from psychology and visual cognition, we find that IAs' judgments are not necessarily influenced by attractiveness. This finding positively reflects one of the profession's main values, namely objectivity. Furthermore, the results for the IA participants underline the importance of professional qualification and training, as the level of qualification, as well as the level of experience, improves the unbiased and objective decision-making and the quality of IAs' judgments.

As real IAs acted as subjects in this study, we assume generalizability and validity of our results. In particular, it seems reasonable to assume that our findings generalize to other internal-audit-related judgments as well as to other professionals' (e.g., external auditors') person-related judgments. However, our study is of course not without limitations.

First, we faced several trade-offs concerning our experimental design, which inevitably entail the aforementioned pros and cons. Second, performance in judgment and decision-making tasks (for further reading see Libby & Luft, 1993), in particular in an (external and internal) auditing setting, might be related to *professional skepticism* (Carpenter, Durtschi, & Gaynor, 2002; Fullerton & Durtschi, 2004; Hurtt, 2010; Nelson, 2009). Though professional skepticism is not the focus of our study, the integration of such aspects into an experiment might be a starting point for further research. Furthermore, future investigations could incorporate judgments of other employees/functions in a company (e.g., management accounting, human resources, or the external auditor). Third, our findings are confined to specific situations: fraud-risk assessments conducted *without direct interaction* with the potential suspect. Mobius and Rosenblat (2006) emphasized that the magnitude of the halo effect may increase in face-to-face-interaction scenarios, which are de facto IAs' routine business. It is therefore possible that objectivity of fraud-risk judgments conducted by experienced IAs might be threatened when direct interactions are involved. Fraud-risk judgments might, however, also benefit from face-to-face interaction because it provides an IA with the possibility to utilize his or her experience and expertise flexibly in interrogations.

Considering these aspects, we suggest that all *initial* fraud-risk judgments should be based on completely neutralized information. If considered necessary, further measures—possibly involving more personal interaction—could be subsequently undertaken. Future research should also determine the effect of judgments performed in teams vs. individual judgements, since typical IAF engagements are performed by more than one IA and audit teams are likely to reduce the probability of individual biases. With this study, we aimed to bridge the gap between the best practice of the IA profession and the scientific knowledge. Our empirical evidence suggests that the objectivity of IAs' judgements may not be affected by physical attractiveness.

ENDNOTES

[1] Fraud risk judgments "may be integrated with an overall organizational risk assessment or performed as a stand-alone exercise, but should, at a minimum, include risk identification, risk likelihood and significance assessment, and risk response" (Institute of Internal Auditors/American Institute of Certified Public Accountants, 2008, p. 7). Inter alia, to evaluate employees' competence and issues of personal integrity can be part of fraud risk judgments (IIA/AICPA, 2008). Though related, and often used in connection with fraud investigations, we refrain from the use of the term forensic auditing in the following, although this notion might be interchangeable.

[2] Internal audit standards define objectivity as "an unbiased mental attitude that allows IAs to perform engagements in such a manner that they believe in their work product and that no quality compromises are made. Objectivity requires that IAs do not subordinate their judgment on audit matters to others" (IIA, 2013, attribute standard 1100). We acknowledge that many potential fraud cases are detected through software and/or screening of documents, which is in particular the case in an

early phase of an audit. However, in later phases, direct personal interaction with the auditee or possible suspects is much more common.

[3] The designs of the main experiments and the surveys used for stimulus creation meet the requirements for using human subjects in the experimental laboratory at the university where the lead author is located. The use of human subjects was also approved by the institution that supported the main experiment with IAs, the German Chapter of the IIA.

[4] As part of the experimental questionnaire, subjects are required to answer the following question on a seven-point Likert scale with endpoints labeled as "very unimportant"/"very important": "In your opinion, how important is it that an IA makes objective judgments?" In line with the expectations, literally all IAs assess objectivity to be "very important."

[5] Note that the model neither requires nor do we expect uniform expertise among IAs; see Gul, Wu, and Yang (2013) and Knechel, Vanstraelen, and Zerni (2015) for differences in individual auditor characteristics.

[6] Potential participants were identified by approaching them during breaks with the support of several doctoral students and were asked if they would be willing to participate in our experiment. Interested individuals were then guided to a separate space at the meeting venue and were randomly presented with one version of the experimental materials (see Section 3.3 for details).

[7] To understand whether participants perceive differences in attractiveness and age of our different stimuli/employee's faces as intended, we asked for their perceptions of *Schmidt's* attractiveness and age. As outlined in Table 5, attractiveness ratings differ significantly at the 1% and 5% significance levels (one-tailed *t*-test) between the IA and naive subject groups respectively. Untabulated analyses also show that age ratings differ significantly at the 1% and 1% significance levels (one-tailed *t*-test) between the IA and naive subject groups respectively. We did not include a manipulation check question for gender, as *Schmidt's* gender is obvious because of the stimuli and the wording of the case (she/he; Ms/Mr).

[8] As can be seen from Table 5, a significant number of subjects in each group answered "no" to the question "Is the risk related to the issue at hand so high that further fraud investigation is necessary?" We reran all reported analyses for those subjects who *do see* the necessity for further fraud investigation only, as their perceptions might differ fundamentally from those who *do not see* the necessity. However, all results remain consistent.

ORCID

Marc Eulerich http://orcid.org/0000-0002-9965-7584
Jochen C. Theis http://orcid.org/0000-0003-3634-5193
Junpeng Lao http://orcid.org/0000-0001-6029-7510
Meike Ramon http://orcid.org/0000-0001-5753-5493

REFERENCES

Abbott, L. J., Park, Y., & Parker, S. (2000). The effects of audit committee activity and independence on corporate fraud. *Managerial Finance*, *26*(11), 55–67.

Adams, G. (1982). *Physical attractiveness*. New York, NY: Praeger.

Alley, T. (1988). Social and applied aspects of face perception: An introduction. In T. Alley (Ed.), *Social and applied aspects of perceiving faces*. Hillsdale, NJ: Lawrence Erlbaum Associates.

American Institute of Certified Public Accountants. (2002). *Considerations of fraud in a financial statement*. New York, NY: AICPA.

Association of Certified Fraud Examiners (2012). *Report to the nations on occupational fraud and abuse. 2012 global fraud study*. Austin, TX: ACFE.

Bashour, M. (2007). *Is an objective measuring system for facial attractiveness possible?* Boca Raton, FL: Universal-Publishers.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beasley, M. S., Carcello, J. V., Hermanson, D. R., & Lapides, P. D. (2000). Fraudulent financial reporting: Consideration of industry traits and corporate governance mechanisms. *Accounting Horizons*, *14*(4), 441–454.

Beasley, M. S., Carcello, J. V., Hermanson, D. R., & Neal, T. L. (2010). *Fraudulent financial reporting: 1998–2007. An analysis of U.S. public companies*. New York, NY: Committee of Sponsoring Organizations of the Treadway Commission.

Berscheid, E. (1986). The question of the importance of physical attractiveness. In C. Herman, M. Zanna, & E. Higgins (Eds.), *Physical appearance, stigma, and social behavior: The Ontario symposium*. Hillsdale, NJ: Erlbaum.

Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype activation and inhibition. In R. S. Wyer, Jr. (Ed.), *Stereotype activation and inhibition: Advances in social cognition*. Mahwah, NJ: Erlbaum.

Carpenter, T., Durtschi, C., & Gaynor, L. (2002). The role of experience in professional skepticism, knowledge acquisition, and fraud detection. Working paper, Florida State University and Georgetown University.

Chaiken, S. (1979). Communicator physical attractiveness and persuasion. *Journal of Personality and Social Psychology*, *37*(8), 1387–1397.

Cooper, D. J., Dacin, T., & Palmer, D. (2013). Fraud in accounting, organizations and society: Extending the boundaries of research. *Accounting, Organizations and Society*, *38*(6–7), 440–457.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285–290.

Dion, K. K. (1973). Young children's stereotyping of facial attractiveness. *Developmental Psychology*, *9*(2), 183–188.

Dion, K. K., & Berscheid, E. (1974). Physical attractiveness and peer perception among children. *Sociometry*, *37*(1), 1–12.

Efran, M. G. (1974). The effect of physical appearance on the judgment of guilt, interpersonal attraction, and severity of recommended punishment in a simulated jury task. *Journal of Research in Personality*, *8*(1), 45–54.

Efran, M. G., & Patterson, E. W. J. (1974). Voters vote beautiful: The effect of physical appearance on a national election. *Canadian Journal of Behavioural Science*, *6*(4), 352–356.

Fanning, K., & Piercey, M. D. (2014). Internal auditors' use of interpersonal likability, arguments, and accounting information in a corporate governance setting. *Accounting, Organizations and Society*, *39*(8), 575–589.

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin*, *111*(2), 304–341.

Fruhen, L., Watkins, C. D., & Jones, B. C. (2015). Perceptions of facial attractiveness, dominance and trustworthiness predict managerial pay awarded in experimental tasks. *Leadership Quarterly*, *26*(6), 1005–1016.

Fullerton, R., & Durtschi, C. (2004). The effect of professional skepticism in the fraud detection skills of internal auditors. Working paper, Utah State University and DePaul University.

Gomulya, D., Wong, E. M., Ormiston, M. E., & Boeker, W. (2017). The role of facial appearance on CEO selection after firm misconduct. *Journal of Applied Psychology*, *102*(4), 617–663.

Gross, A. E., & Crofton, C. (1977). What is good is beautiful. *Sociometry*, *40*(1), 85–90.

Gul, F. A., Wu, D., & Yang, Z. (2013). Do individual auditors affect audit quality? Evidence from archival data. *The Accounting Review*, *88*(6), 1993–2023.

Hart, W., Ottati, V. C., & Krumdick, N. D. (2011). Physical attractiveness and candidate evaluation: A model of correction. *Political Psychology*, *32*(2), 181–203.

Horai, J., Naccari, N., & Fatoullah, E. (1974). The Effects of expertise and physical attractiveness upon opinion agreement and liking. *Sociometry*, *37*(4), 601–606.

Hurtt, R. K. (2010). Development of a scale to measure professional skepticism. *Auditing: A Journal of Practice & Theory*, *29*(1), 149–171.

Institute of Internal Auditors (2013). *International standards for the professional practice of internal auditing*. Altamonte Springs, FL: The Institute of Internal Auditors.

Institute of Internal Auditors/American Institute of Certified Public Accountants. (2008). Executive summary. In *Managing the business risk of fraud: A practical guide*. Retrieved from https://www.aicpa.

org/forthepublic/auditcommitteeeffectiveness/auditcommitteebrief/downloadabledocuments/managing_the_business_risk_of_fraud.pdf

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(19), 7241–7244.

Jackson, L. (1992). *Physical appearance and gender: Sociobiological and sociocultural perspectives*. Albany, NY: State University of New York Press.

Joseph, W. B. (1982). The credibility of physically attractive communicators: A review. *Journal of Advertising*, *11*(3), 15–24.

Kaplan, S. E., Pope, K. R., & Samuels, J. A. (2010). The effect of social confrontation on individuals' intentions to internally report fraud. *Behavioral Research in Accounting*, *22*(2), 51–67.

Kasper-Fuehrera, E. C., & Ashkanasy, N. M. (2001). Communicating trustworthiness and building trust in interorganizational virtual organizations. *Journal of Management*, *27*(3), 235–254.

Knechel, W. R., Vanstraelen, A., & Zerni, M. (2015). Does the identity of engagement partners matter? An analysis of audit partner reporting decisions. *Contemporary Accounting Research*, *32*(4), 1443–1478.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*(3), 390–423.

Leventhal, G., & Krate, R. (1977). Physical attractiveness and severity of sentencing. *Psychological Reports*, *40*, 315–318.

Libby, R., & Luft, J. (1993). Determinants of judgment performance in accounting settings: Ability, knowledge, motivation, and environment. *Accounting, Organizations and Society*, *18*(5), 425–450.

Little, A. C. (2014). Facial appearance and leader choice in different contexts: Evidence for task contingent selection based on implicit and learned face-behaviour/face-ability associations. *The Leadership Quarterly*, *25*(5), 865–874.

Lorenzo, G. L., Biesanz, J. C., & Human, L. J. (2010). What is beautiful is good and more accurately understood: Physical attractiveness and accuracy in first impressions of personality. *Psychological Science*, *21*(12), 1777–1782.

Marlowe, C. M., Schneider, S. L., & Nelson, C. E. (1996). Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased? *Journal of Applied Psychology*, *81*(1), 11–21.

Mazzella, R., & Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta-analysis. *Journal of Applied Social Psychology*, *24*, 1315–1344.

Michelini, R. L., & Snodgrass, S. R. (1980). Defendant characteristics and juridic decisions. *Journal of Research in Personality*, *14*(3), 340–350.

Miller, A. G. (1970a). Role of physical attractiveness in impression formation. *Psychonomic Science*, *19*(4), 241–243.

Miller, A. G. (1970b). Social perception of internal–external control. *Perceptual and Motor Skills*, *30*(1), 103–110.

Mobius, M. M., & Rosenblat, T. (2006). Why beauty matters. *American Economic Review*, *96*(1), 222–235.

Moeller, R. (2004). *Sarbanes–Oxley and the new internal auditing rules*. Hoboken, NJ: John Wiley & Sons, Inc.

Mutchler, J. (2003). Independence and objectivity: A framework for research opportunities in internal auditing. In The Institute of Internal Auditors Research Foundation (Ed.), *Research opportunities in internal auditing*. Altamonte Springs, FL: The Institute of Internal Auditors.

Nelson, M. W. (2009). A model and literature review of professional skepticism in auditing. *Auditing: A Journal of Practice & Theory*, *28*(2), 1–34.

Norman, C. S., Rose, A. M., & Rose, J. M. (2010). Internal audit reporting lines, fraud risk decomposition, and assessments of fraud risk. *Accounting, Organizations and Society*, *35*(5), 546–557.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States*, *105*(32), 11087–11092.

Patzer, G. (2006). *The power and paradox of physical attractiveness*. Boca Raton, FL: Brown Walker Press.

Prawitt, D. F., Sharp, N. Y., & Wood, D. A. (2012). Internal audit outsourcing and the risk of misleading or fraudulent financial reporting: Did Sarbanes–Oxley get it wrong? *Contemporary Accounting Research*, *29*, 1109–1136.

Przygodzki-Lionet, N., Olivier, J., & Desrumaux, P. (2010). The effects of facial attractiveness, sex, internality of applicants on hiring decisions for managerial and non-managerial jobs. *Studia Psychologica*, *52*(1), 53–57.

Quereshi, M. Y., & Kay, J. P. (1986). Physical Attractiveness, age, and sex as determinants of reactions to resumes. *Social Behavior and Personality: An International Journal*, *14*(1), 103–112.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Rahn, W. M., & Cramer, K. M. (1996). Activation and application of political party stereotypes: The role of television. *Political Communication*, *13*, 195–212.

Ramon, M., Miellet, S., Dzieciol, A. M., Konrad, B. N., Dresler, M., & Caldara, R. (2016). Super-memorizers are not super-recognizers. *PLoS One*, *11*(3), e0150972.

Richoz, A. R., Jack, R. E., Garrod, O. G. B., Schyns, P. G., & Caldara, R. (2015). Reconstructing dynamic mental models of facial expressions in prosopagnosia reveals distinct representations for identity and expression. *Cortex*, *65*, 50–64.

Rohner, J. C., & Rasmussen, A. (2011). Physical attractiveness stereotype and memory. *Scandinavian Journal of Psychology*, *52*, 309–319.

Rule, N., & Tskhay, K. O. (2014). The influence of economic context on the relationship between chief executive officer facial appearance and company profits. *The Leadership Quarterly*, *25*, 846–854.

Shinners, E. (2009). *Effects of the "what is beautiful is good" stereotype on perceived trustworthiness*. (pp. 1–5). XII: University of Wisconsin–La Crosse Journal of Undergraduate Research.

Sigall, H., & Ostrove, N. (1975). Beautiful but dangerous: Effects of offender attractiveness and nature of the crime on juridic judgment. *Journal of Personality and Social Psychology*, *31*, 410–414.

Snyder, H., Clifton, J., & Bowlin, W. (2012). Alchemy—An internal auditing case. *IMA Educational Case Journal*, *5*(4). article 2

Snyder, M., & Rothbart, M. (1971). Communicator attractiveness and opinion change. *Canadian Journal of the Behavioural Science*, *3*, 377–387.

Szulanski, G., Cappetta, R., & Jensen, R. J. (2004). When and how trustworthiness matters: Knowledge transfer and the moderating effect of causal ambiguity. *Organization Science*, *15*, 600–613.

Torrance, J. S., Wincenciak, J., Hahn, A. C., DeBruine, L. M., & Jones, B. C. (2014). The relative contributions of facial shape and surface information to perceptions of attractiveness and dominance. *PLoS One*, *9*(10), e104415.

Von Stockhausen, L., Koeser, S., & Sczesny, S. (2013). The gender typicality of faces and its impact on visual processing and on hiring decisions. *Experimental Psychology*, *60*(6), 444–452.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. *Advances in Experimental Social Psychology*, *29*, 141–208.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer-Verlag.

Yu, H., Garrod, O. G. B., Jack, R., & Schyns, P. G. (2015). A framework for automatic and perceptually valid facial expression generation. *Multimedia Tools and Applications*, *74*(21), 9427–9447.

Yu, H., Garrod, O. G. B., & Schyns, P. G. (2012). Perception-driven facial expression synthesis. *Computers & Graphics*, *36*(3), 152–162.

Zimbelman, M. F. (1997). The effects of SAS No. 82 on auditors' attention to fraud risk factors and audit planning decisions. *Journal of Accounting Research*, *35*(Supplement), 75–97.

Marc Eulerich holds the Chair for Internal Auditing at the University of Duisburg–Essen and has a strong research focus on internal auditing in the corporate governance framework. After some practical experience, he has held the professorship since 2011.

Jochen Theis, after finishing his Business Administration Degree at the University of Mannheim, earned his PhD at the Chair for Accounting, Auditing and Controlling at the University of Duisburg–Essen and is now a post-doc researcher with a focus on experimental research in auditing.

Junpeng Lao is working as a post-doc researcher at the University of Fribourg, Department of Psychology. He is interested in computation and mathematics and focuses on the computational nature of neural systems using various techniques.

Meike Ramon is working as a post-doc at the University of Fribourg with a focus on the effects of experience on visual perception, both in healthy and neuropsychological populations. She has numerous publications and interdisciplinary research projects.

How to cite this article: Eulerich M, Theis JC, Lao J, Ramon M. Do fine feathers make a fine bird? The influence of attractiveness on fraud-risk judgments by internal auditors. *Int J Audit*. 2018;1–13. https://doi.org/10.1111/ijau.12137

APPENDIX

EXPERIMENTAL MATERIALS AND QUESTIONNAIRE

[Format differs from original questionnaire].

Please read the following information carefully!

ALCHEMY—AN INTERNAL AUDITING CASE

Introduction to Auditing Alchemy

Auditing Alchemy (AA) is a privately held manufacturing company established in 1998. AA manufactures spheres that are used in a variety of industrial applications. AA is the top sphere producer in the industry; the company sells spheres to the US government and holds other aerospace contracts. The spheres are sold in two forms: green and gold. A third form of sphere, red, is a by-product of the manufacturing process.

Although it is not publicly traded, AA has had an internal audit department for more than a decade. The department has a chief audit executive (CAE) who reports to the board of directors and who administratively reports to the CEO. The department has three other internal auditors in addition to the director. The department has board responsibilities within AA, but among its duties are deterring and investigating fraud and safeguarding assets.

Please read the following information and dialogue carefully!

Furthermore, please assume that you take on the role of the lead auditor Mr. Meier.

Further information

Sylvia Torrance is Auditing Alchemy's CEO. She has been feeling uneasy about AA's manufacturing output since the beginning of calendar year 2006. Although sales were brisk and profits remained at an acceptable level, there seemed to be less profit from sales and higher expenses for manufacturing. As a result, she scheduled a meeting with AA's CAE Martin Meier to discuss the situation.

Conversation between Sylvia Torrance (CEO) and Martin Meier:

ST: Thanks for stopping by today. I'm not sure if we have a problem here, but something doesn't seem right about our finances. The profit level is acceptable, but we should be doing better. Our gross margin percentage dropped from 27% in 2005 to 25% so far in 2006. Our selling price and operating expenses are pretty consistent between the two years.

MM: What is it exactly that doesn't seem right to you?

ST: That's just the problem. I can't quite put my finger on it, but the financials for the company just don't look right to me. The bottom line is still healthy, but I'm just not comfortable with things. The CFO's staff has done some preliminary analyses, and it appears there might be a problem with missing inventory. Let me lay out for you what we have found so far. First, there was an extensive examination of our warehouse. The physical protection for inventory is the most secure I've ever seen. We also did a complete review of the internal controls; they're well designed and, more importantly, they actually work the way they are intended. I am confident that if there is a problem with missing inventory, it has occurred in manufacturing, before it reaches the warehouse. Our raw material cost per unit to produce a gold sphere and a green sphere is up significantly over the last couple of years. We use a standard cost system and consequently develop a standard material cost for each type of sphere produced. Our standard material cost is EUR 348 for a gold sphere and EUR 90 for a green sphere. Last year, our costs were just about on target (EUR 349 for gold spheres and EUR 89 for green spheres) compared with the standard but not this year. This year, the material cost for the gold spheres has increased 2% to EUR 356. We further broke the materials cost variance into a price variance and a usage variance. The price variance was near zero, as I had expected because purchasing had told me there was no increase in the price we were paying for raw materials. In fact, the whole cost variance was made up of the usage variance. We used more raw material in producing this year's usable spheres than we had in previous years. At the same time, however, our cost of goods manufactured/raw material inventory ratio is pretty consistent over the most recent six quarters, indicating that we are using about the same amount of raw material in our production process as we have previously. We use a perpetual inventory system and have not done a manual inventory count for years because of the cost of doing so. Finally, we have also noticed that our raw material purchases are up over last year. I know that this is a hodge-podge of information, but it all points to us using more materials than we should. I've talked with the

production manager, Schmidt, about whether *she/he* can identify any reason why we would be using more materials, such as labor not being sufficiently trained on the new machinery we bought, raw materials being of poor quality, or machinery not having the preventive maintenance being done. However, *she/he* said none of those factors are a problem. I'm at a loss as to what more I can do, so I'm seeing if you might be able to help. I have attached some information for you about Mr./Ms. Schmidt. Will you look into it?

MM: Of course. Do you suspect fraud?

ST: I don't know. I'm not sure what the problem is, but we don't seem to have the money and profits we ought to have, and the market is booming. What else could it be?

MM: I'm not sure, but before we start shaking things up with a word like "fraud," I should look into things in more detail. A full-blown fraud examination costs a bundle and it puts everybody on edge, even if they're honest. Tell you what. I'll run some quick diagnostics and see if anything falls out, and then we can decide whether things are as bad as you think they are.

Please answer the following questions.

Furthermore, please assume that you take on the role of the CAE, Mr. Meier.

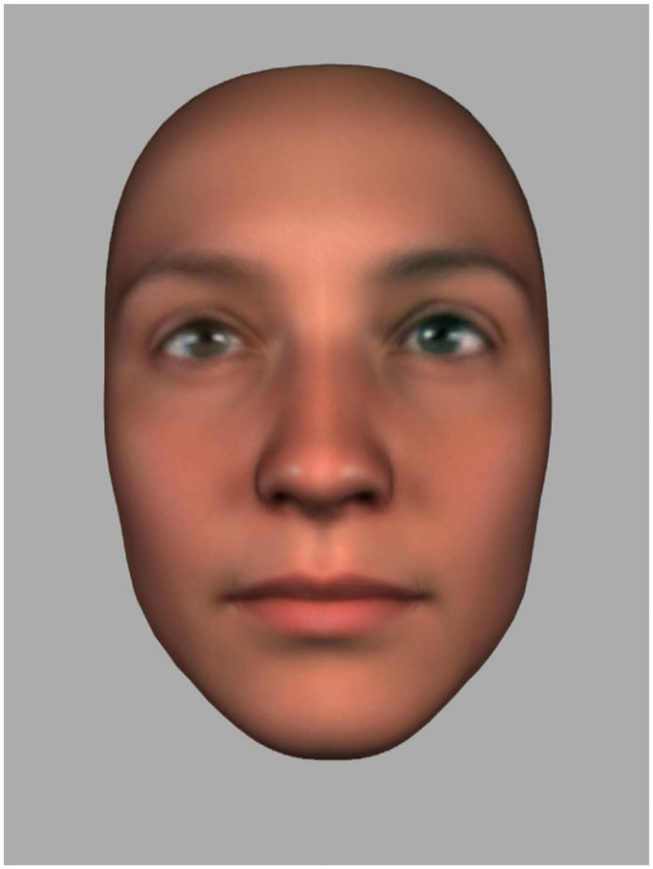
Is the risk related to the issue at hand so high that further fraud investigation is necessary?								
□ Yes				□ No				
How high is the risk related to the issue at hand?								
Very low	−3 □	−2 □	−1 □	0 □	1 □	2 □	3 □	Very high
How do you assess the probability that Mr./Ms. Schmidt is responsible for the issue/possible fraud?								
Very improbable	−3 □	−2 □	−1 □	0 □	1 □	2 □	3 □	Very probable
In your opinion, how important is it that an internal auditor makes objective judgments?								
Not very important	−3 □	−2 □	−1 □	0 □	1 □	2 □	3 □	Very important
How attractive do you find Mr./Ms. Schmidt?								
Not very attractive	−3 □	−2 □	−1 □	0 □	1 □	2 □	3 □	Very attractive
How old is Mr./Ms. Schmidt?								
Very young	−3 □	−2 □	−1 □	0 □	1 □	2 □	3 □	Very old

Please also answer the following questions:

Your work experience overall (in completed years)?	✍ _____
Your work experience in internal auditing (in completed years)?	✍ _____
The number of your full-time subordinates (0 if you have no subordinates)?	✍ _____
What is your highest educational degree? (e.g., Bachelor's, Diploma, Master's, PhD)?	✍ _____
Please list all other relevant qualifications or certifications you possess (e.g., CPA, CIA, etc.).	✍ _____
What is your present professional position (e.g., auditor, senior auditor, team leader, department head, division manager, lead auditor)?	✍ _____
Please tell us your age.	✍ _____
I am…	❑ female ❑ male

[Example: Manipulation YFA (young/female/attractive). The experimental materials were arranged in a folder to ensure that the extract from the personnel file was constantly visible to the participants during their work on the case.]

Extract from the personnel file of the production manager Ms. Schmidt



Personal Information	**Employee**
Name:	Schmidt, Marion
Duration of employment:	5 years
Existing entries in the personnel file	none
Address:	[Street anonymized for review] [City anonymized for review], Germany Phone: (+49) [Number anonymized for review]
Account details	XYZ Bank
Code of the payment service provider	365 XXX XX
Bank account number	378 XXXX