Routledge
Taylor & Francis Group

Journal of Behavioral Finance

ISSN: 1542-7560 (Print) 1542-7579 (Online) Journal homepage: http://www.tandfonline.com/loi/hbhf20

Investment Professionals' Ability to Detect Deception: Accuracy, Bias and Metacognitive Realism

Maria Hartwig, Jason A. Voss, Laure Brimbal & D. Brian Wallace

To cite this article: Maria Hartwig, Jason A. Voss, Laure Brimbal & D. Brian Wallace (2017) Investment Professionals' Ability to Detect Deception: Accuracy, Bias and Metacognitive Realism, Journal of Behavioral Finance, 18:1, 1-13

To link to this article: http://dx.doi.org/10.1080/15427560.2017.1276069

JOURNAL OF BEHAVIORAL FINANCE
2017, VOL. 18, NO. 1, 1–13
http://dx.doi.org/10.1080/15427560.2017.1276069

Routledge
Taylor & Francis Group

Investment Professionals' Ability to Detect Deception: Accuracy, Bias and Metacognitive Realism

Maria Hartwig[a], Jason A. Voss[b], Laure Brimbal[a], and D. Brian Wallace[a]

[a]City University of New York; [b]CFA Institute

ABSTRACT

In the first empirical study on the topic, the authors examined the ability of investment professionals to distinguish between truthful and deceptive statements. A random sample of 154 investment professionals made judgments about a series of truthful and deceptive statements, some of which involved financial fraud. Investment professionals' lie detection accuracy was poor; participants performed no better than would be expected by chance. Accuracy in identifying lies about financial fraud was especially poor. Further, participants displayed poor metacognitive realism when assessing their own performance. The theoretical and practical implications for lie detection in the financial industry are discussed.

KEYWORDS

Interpersonal deception; Investment professionals; Lie detection; Metacognitive realism

Introduction

Common morality holds that lying is bad (Anderson [1968], Backbier, Hoogstraten, and Meerum Terwogt-Kouwenhoven [1997]). Despite this view, research shows that people lie frequently in everyday life (DePaulo et al. [1996], DePaulo and Kashy [1998]). In some contexts, being able to detect lies is critical. Examples of such contexts are the legal system, national security, and intelligence collection. A vast body of research has examined deception and its detection in these settings. However, there is to date little research on lie judgments in financial analysis settings, despite large monetary incentives to deceive on the part of executives, and also a long history of such deceptions. Furthermore, the consequences of failing to detect deception by financial analysts are sizable, and often amount to billions of dollars in losses for investors in deceptive businesses. In part, deception occurs because of performance-based compensation plans for corporate executives whose results depend on the subjective accounting judgments of those same professionals. These conflicts of interest and monetary incentives to manipulate information lead to fraudulent behavior that is largely undetectable, except in evaluating the veracity of executive statements about the performance of their companies. In short, there are opportunities for deception with high stakes for would-be liars and for the financial professionals trying to detect that deception.

This study examined fundamental questions about the characteristics of investment professionals' judgments of deception. We did this by showing recordings of deceptive and truthful statements told in different contexts (including financial settings) to a sample of investment professionals, and asking them to provide their judgment of the speakers and of their own performance as lie detectors. In order to develop a background for this study, we provide an overview of the literature on human lie judgments, with a particular focus on people's ability to distinguish between true and false statements. Subsequently, we will briefly discuss deception in the financial realm. We then describe what is to our knowledge, the first empirical examination of investment professionals' ability to detect deception.

Lying has been of interest to scholars for centuries. There is now a substantial body of scientific research on deception spanning half a century (Vrij [2008]). Some of this work focuses on the dynamics of deception in everyday life. Studies show that people report lying on an everyday basis, often with little hesitation (DePaulo et al. [1996], but see Serota and Levine [2015]). People lie to strangers, but they also lie to those with whom they have close relationships (DePaulo, Epstein, and Wyer [1993], DePaulo and Kashy [1998]). Thus, social psychological research shows that lying is a common feature of ordinary social life, and that people tend to lie with relative

CONTACT Jason A. Voss jason.voss@cfainstitute.org 3703 Kingston Blvd., Sarasota, FL 34238.

Color versions of one or more figures in the article can be found online at www.tandfonline.com/hbhf.

ease and little regret. We shall return to the consequences of these findings later in this article.

Cues to deception

A major theme in the deception literature has been the search for cues to deception—behavioral indicators that distinguish between truths and lies. The ultimate cue to deception would be a metaphorical Pinocchio's nose; a behavior or set of behaviors that always occurs when people are lying, but never when they are telling the truth. In order to examine whether such a phenomenon exists, researchers have conducted extensive comparisons of the behavioral differences between liars and truth tellers.

The soundest way to grasp the findings from a vast body of literature is through meta-analyses, which involve statistical synthesis of an entire realm of research. The most comprehensive meta-analysis to date was conducted by DePaulo et al. [2003], and included analyses of 158 behaviors. These included nonverbal cues such as fidgeting, eye movements, posture changes; verbal cues such as quantity and quality of details included in the statement; and paralinguistic cues such as speech errors, hesitations, and pauses. The results of this massive analysis are easy to summarize: Only a small number of behaviors appear to be related to deception. For example, liars appear less cooperative, more tense and ambivalent, and their stories are less compelling. However, these cues are relatively weak in their relationship with deception. The conclusion from research on cues to deception is thus that the notion of a Pinocchio's nose is fictional.

Human lie judgments

How accurate are people at distinguishing between true and false statements? This is a fundamental question in the deception literature. Researchers typically examine this question by showing participants statements that are either true or false, and asking them to make a judgment about whether the statement is a truth or a lie. Simply guessing would yield an accuracy rate of 50%. A major meta-analysis of the deception literature including over 200 samples showed an average accuracy rate of 54%, which is only marginally higher than chance performance (Bond and DePaulo [2006]). Moreover, the data show that people typically display a truth bias—a tendency to judge statements as true rather than false (Bond and DePaulo [2006]). Thus our first two research questions are: First, with what accuracy are investment professionals able to distinguish between true and false statements? Second, do they display a judgment bias?

Accuracy of lie judgments in experts

Perhaps the reader wonders whether accuracy rates are higher under certain circumstances. That is, are there variables moderating the accuracy rate of human judgments? Contrary to common beliefs, the above-chance accuracy is remarkably stable. Even presumed lie experts, that is, people who have to make judgments of credibility as part of their professional lives (e.g., law enforcement officers, customs officers, judges, lawyers, migration agencies) obtain accuracy rates around chance levels (Vrij [2008]). These presumed experts are not identical to lay people in their decision making about deception, though: Laypeople are prone to a truth bias—but in contrast, professional lie catchers instead display a chronic tendency toward suspicion (Meissner and Kassin [2002]). They are also more overconfident than lay people (Garrido, Masip, and Herrero [2004], Kassin et al. [2007]). Given these findings, the literature does not provide the basis for optimism about the performance of investment professionals; we predict an accuracy rate only slightly higher than chance. As for predictions about judgment bias, we abstain from making specific predictions as the literature is mixed: While some professional groups have been found to display a lie bias (Meissner and Kassin [2002]), the literature in general shows that people tend toward a truth bias (Bond and DePaulo [2006]).

Explanations for poor accuracy rates

Why do people obtain such mediocre hit rates? Two primary explanations have been offered. The first is that the cues that people believe to be related to deception (so called subjective cues) do not overlap with actual, objective cues to deception. This explanation has been labeled the wrong subjective cue hypothesis. There is a substantial literature mapping people's beliefs about the characteristics of deceptive behavior (Hartwig and Granhag [2015], Strömwall, Granhag, and Hartwig [2004]). This literature covers a range of groups and cultures (e.g., laypeople, police officers, criminals, business managers). The results are remarkably consistent: People subscribe to a universal stereotype of deceptive behavior. More specifically, people believe that liars avert their gaze and that they display nervous and fidgety behavior (Global Deception Research Team [2006]). Of particular relevance for this paper, a recent survey examined investment professionals' beliefs about deceptive behavior (Hartwig, Voss, and Wallace [2015]). It showed that investment professionals share the widespread belief that liars betray themselves through gaze aversion and fidgeting. It also showed that investment professionals have

some faith in their ability to detect deception—their self-estimated accuracy rates exceeded 65% (which is considerably higher than the 54% hit rate found in meta-analyses).

Comparisons between people's beliefs about deception and the literature on actual cues to deception show that people indeed have misconceptions about deceptive behavior (Strömwall et al. [2004]). Gaze aversion is not a valid indicator of lying, nor is nervous behavior and fidgeting (DePaulo et al. [2003]).

Another explanation for poor accuracy in lie judgment is the inherent difficulty of the task. That is, it may be that lie catchers fail because the behavioral differences between liars and truth tellers are so miniscule. A large-scale meta-analysis recently tested this explanation, and found that the scarcity of reliable cues to deception is indeed a major contributor to poor performance. In short, even if lie catchers were disabused of false stereotypes of liars, they would not achieve a substantially higher hit rate (Hartwig and Bond [2011]).

Consequences of poor lie detection accuracy

In social life, lies are typically relatively trivial. Failure to detect these lies therefore usually does not carry severe penalties. In other settings however, being able to distinguish between truthful and deceptive statements is of critical importance. In the literature on deception, most attention has been focused on legal and security settings. Much less attention has been directed to deception and its detection in the financial industry. This lack of research is both surprising and problematic, as judgment of truth and deception are of great importance in this domain given the long history of business fraud. A small sampling of recent market capitalization losses borne by equity investors due to fraud includes: Australia's Qintex's 1989 losses of AU$1.5 billion; Adelphia Communications' 2002 losses of $8.4 billion; Enron's 2001–2002 losses of over $70 billion; Worldcom's 2001–2002 losses of $186 billion; Nortel's 2000–2009 losses of $283 billion; Parmalat's 2003 losses of €3.7 billion; and Banco Espirito Santo's 2014 losses of €95 billion. Losses to debt holders due to fraud are more difficult to estimate because many recoup a portion of their principal investment through bankruptcy litigation. However, it is safe to assume the total is well into the billions for this asset class, too. In each of the examples cited reported financial results were not just misstated, but also fraudulent, escaping the notice even of these firms' auditors. In order to spot these frauds financial analysts would therefore have to rely on the accuracy of their judgments about the veracity of corporate executive statements about financial performance. In other words, financial analysts must be lie catchers, in addition to accounting experts, in order to do their jobs well.

It should be noted that lie catchers can commit two different type of errors: They can mistakenly judge a true statement as being deceptive (e.g., a so called false positive error), or they can mistakenly believe that a deceptive statement is truthful (e.g., a so called false negative error). The consequences of these errors depend on the setting. In financial settings, analysts committing a false positive error believe a company is engaged in fraud, or that an executive has made a deceptive statement, yet the firm, the executive, or both are speaking truthfully. Here, if the financial analyst does not purchase shares in the business, or sells shares of stock already owned, and the price of the shares subsequently rises the error results in opportunity costs. Assessing the financial consequences in the case of a false positive is difficult, because the financial analyst may invest foregone monies in other opportunities that subsequently rise. However, at the very least the financial analyst is incorrect in his or her assessment of the company and its executives. Consequences of a false negative are much easier to estimate as these are the losses in asset value due to fraud.

Metacognitive realism

As part of judgments about veracity, people often make metacognitive judgments. Metacognition broadly refers to awareness and assessments of one's own thoughts and judgments (Schwarz [2015]). Here, we focus on one of the most common metacognitive judgments that accompany human decision making: confidence judgments. Confidence judgments refer to the degree of certainty one has that one's judgment is correct. We constantly make (and request others to provide) confidence judgments about a variety of psychological processes, including memory (e.g., how certain are you that this event happened?) and interpersonal impression (e.g., how certain are you that this person is trustworthy?). Our third research question concerns investment professionals' level of metacognitive realism when it comes to their own lie detection ability.

Metajudgments of confidence may serve to regulate the extent to which one operates on one's judgments. For example, a person who is strongly confident in his assessment that his partner is cheating on him is more likely to pursue the issue than a person whose confidence is more tempered. A key question in the study of metacognition is the relationship between confidence and accuracy. If the relationship between confidence and accuracy is weak, reliance on confidence as an indicator of accuracy is misguided. A large body of work has examined the realism (or calibration) of people's

confidence judgments. This literature shows that people have a general tendency toward overconfidence (Lichtenstein and Fischhoff [1977]).

Metacognitive realism has been examined in relation to judgments of deception. A meta-analysis of the deception literature found that the accuracy-confidence correlation in deception judgment was indistinguishable from zero (DePaulo et al. [1997]). This means that there is no reason to think that a person who claims to be 100% confident in their judgment about veracity will be more accurate than a person who claims to be only 50% accurate (which amounts to guessing). Our last hypothesis regards metacognitive realism of investment professionals about their lie detection ability. We predict two broad findings based on the previous literature: First, we expect investment professionals to display poor calibration—that is, we expect that confidence will be poorly related to accuracy. Second, we predict that investment professionals will display overconfidence in their ability to judge deception.

Present study

In this study, our focus is on the accuracy, judgment bias, and metacognitive realism of investment professionals. As discussed previously, deception judgments may play an important role for investment professionals, yet there is no systematic data on their performance. To our knowledge, the only study of investment professionals and deception was the survey by Hartwig, Voss, and Wallace [2015]. We aim to address this gap in the literature by examining investment professionals' judgments of deception, based on a broad set of stimulus materials including laboratory and real-life, high-stakes lies and truths. In order to provide an externally valid test of investment professionals' performance in their domain of work, we included a sample of fraudulent and truthful statements made during quarterly earnings calls.

Method

Participants

Participants ($N = 433$) were finance professionals, affiliated with CFA Institute at the time of the study. Of the 433 participants who followed the link to the study, our final sample involved 215 who completed at least one lie-truth judgment. The sample was predominantly composed of men ($n = 142$, 91% of those reporting gender) between the ages of 24 and 68 years old ($M = 39.03$ years, $SD = 8.95$ years; see Figure 1 for distribution). Participants reported varied levels of experience, with 10–20 years being the most common ($n = 67$, 43%), followed by 0–10 years ($n = 59$) and 20+ years ($n = 30$). Participants reported a wide range of professions, mostly financial analysts, and advisors and asset managers (for more details, see Figure 2). These demographics indicate that our sample was fairly representative of the finance industry as a whole (e.g., http://www.eeoc.gov/eeoc/statistics/reports/finance/index.html) (U.S. Equal Employment Opportunity Commission [2006]).

Procedure

Participants were recruited randomly by the CFA Institute. All participants received an email asking them to participate in the present study. The email contained a

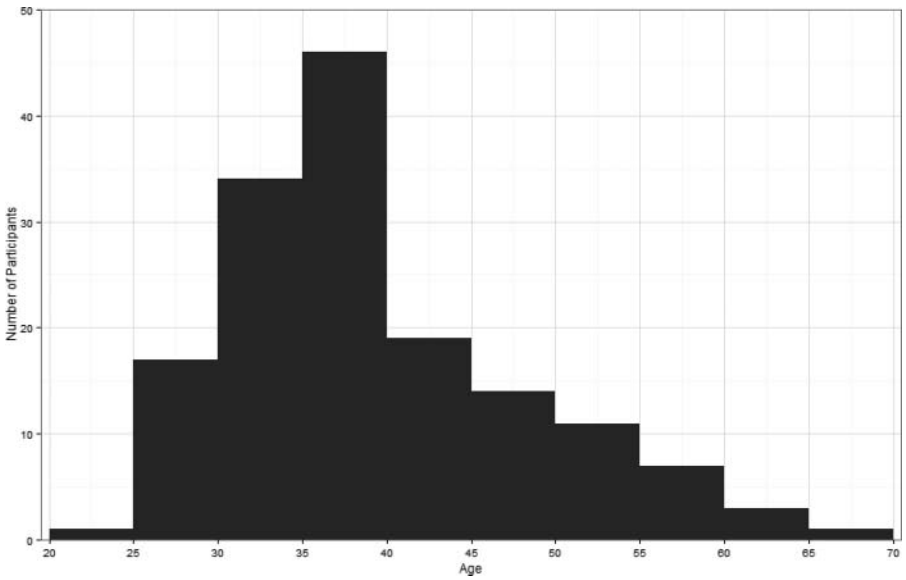

Figure 1. Participants' age distribution.

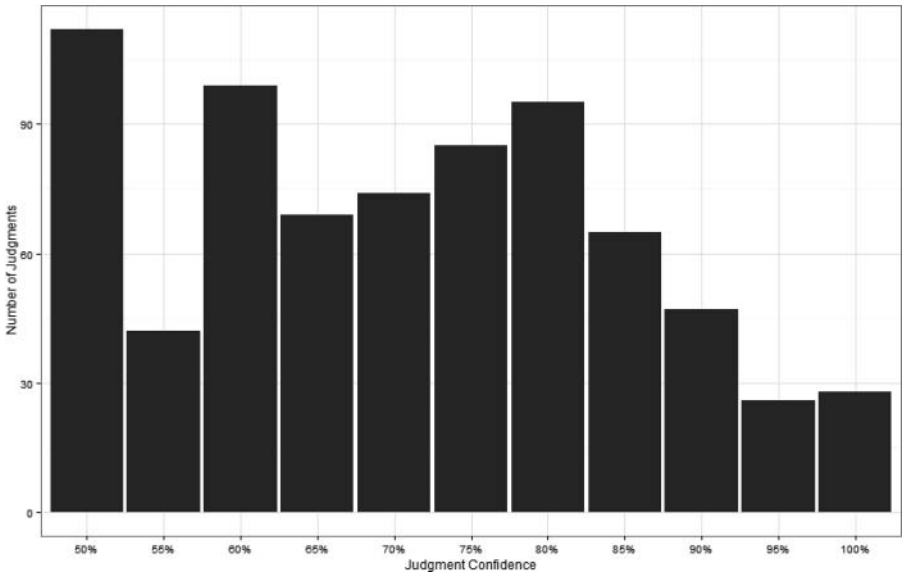

Figure 2. Number of judgments made at each level of confidence.

link that redirected interested individuals to an online system housing the study. Once informed consent was obtained, participants were presented with three videos and one audio recording of statements and were asked to determine the veracity of each one. We counterbalanced the order of presentation of statements and we had several versions of each type that were randomly selected for each participant. This means that one participant would not necessarily see any of the same videos or in the same order as another participant. After each statement, participants were given the dependent measures. Finally, a short set of demographic information was collected.

Materials

To achieve a large and representative sample of lies and truths, we used four sets of stimulus material. Three of them were videos, created and provided by the authors of previous interviewing and deception detection research (see the following examples). The fourth sample of stimuli was a set of audio recordings of publicly traded U.S. companies' financial quarterly reports, provided by one of the authors. Participants were given a short contextual description prior to exposure to each type of statement so as to provide them with an understanding of the speaker's situation.

Video set 1 (Sorochinski et al. [2014])

The statements from this study were based on interviews with mock suspects, who either lied or told the truth about their involvement in a staged act of terrorism. We randomly sampled from the 116 videos to obtain a subsample of 20 videos (10 lies and 10 truths)

Video set 2 (Toomey [2013])

Participants in this study were convicted felons who, similarly to the participants in Sorochinski et al. [2014], either lied or told the truth about a mock crime. From the total sample of 70 videos, we randomly selected 10 truths and 10 lies.

Video set 3 (Vrij and Mann [2001])

This study employed videos of real-life, high stakes lies. They were tapes of press conferences depicting people either pleading for the return of a close one who had gone missing or asking for information that could help identify the murder of a relative whose body had been found. Five of these pleaders were subsequently found to be implicated in the disappearance or murder of their loved one and thus were being deceptive during the press conference. Three of the pleaders were cleared of guilt and thus considered truthful during their press conference.[1] Vrij and Mann [2001] provided a total of eight videos (five lies, three truths), all of which we used.

Financial audio statements

All of the audio statements were taken from conference call recordings cross-referenced between U.S. Securities and Exchange Commission (SEC) Enforcement Actions and an archive of conference call recordings owned by Bloomberg. These recordings feature companies (IndyMac Bancorp, Hansen Medical, and Office Depot) before they were exposed to have hidden or withheld information that lead to their prosecution via SEC Enforcement Action. Each one of these cases lead the company in question to lose credibility as well as tremendous value as measured by declines in market capitalization. These recordings were specifically matched to

language identified as false or misleading by the SEC in its Enforcement Actions. Each identified statement was considered a lie for purposes of creating the stimulus materials. As every statement from the quarterly calls was not deceptive, it was possible to create truthful statements from the same quarterly calls. For a more detailed description of each call, see the Appendix.

Dependent measures

After each statement, participants were asked to respond to several questions. The first was a dichotomous guilt versus innocence rating, followed by a rating of how confident they were in their assessment. The third question was a quantitative measure of the truthfulness of the statement. In addition to these measures, after the financial statements, participants were asked if they recognized from which company the call originated and if they were working on this call. This was to ensure participants were not bringing their own knowledge of the cases into their veracity assessments.

Demographics

After responding to the dependent measures for all videos, participants were given a short demographic questionnaire. Participants were asked to provide their age (open-ended response), profession (open-ended response), gender (male, female, or other), and years of experience (0–2 years, 2–5 years, 5–10 years, 10–20 years, 20–30 years, or more than 30 years).

Results

Of the 443 participants who followed the link and viewed the informed consent page, just under half ($n = 215$) evaluated at least one video. Of those, most ($n = 154$) completed the experiment by providing judgments on each of the four video stimuli types. The participants who did not finish dropped out after evaluating one ($n = 23$), two ($n = 21$), or three ($n = 17$) videos.

Statistical analysis has been moving away from Null Hypothesis Significance Testing (NHST) and its results section incarnation, the p value, for some time (Carver [1978]). We have opted to follow this trend, and to focus on reporting confidence intervals and effects sizes of interest. Instead of p values, any NHST performed here will be based on bootstrapped bias-corrected and accelerated (BCa) confidence intervals.

Lie detection accuracy

On the dichotomous judgment of truth or lie, participants made a correct decision 49.4% of the time—an accuracy rate that did not significantly differ from merely guessing (50%), $M = 0.494$, 95% CI [0.447, 0.540]. Performance was best for the Sorochinski et al. [2014] videos ($M = 0.548$, 95% CI [0.472, 0.623]), and worst for the Toomey [2013] videos ($M = 0.478$, 95% CI [0.406, 0.552]). Accuracy in evaluating finance statements (51.8%) was slightly better, but not significantly so, than accuracy in nonfinance statements (50.2%)—only a 1.6% difference ($M_{diff} = 0.016$, 95% CI [−0.066, 0.097]).

The continuous judgment of truthfulness revealed a similar pattern. No significant overall difference between truthful stimuli ($M = 5.32$, 95% CI [5.07, 5.56]) and deceptive stimuli ($M = 5.50$, 95% CI [5.25, 5.74]), was observed, $M_{diff} = -0.17$, 95% CI [−0.52, 0.17], $d = -0.07$. Only one set of stimuli showed significant differences in mean truthfulness between the two veracities—participants rated truthful Vrij and Mann [2001] videos ($M = 3.87$, 95% CI [3.30, 4.50]) as less truthful than deceptive ones ($M = 4.64$, 95% CI [4.26, 5.05]), $M_{diff} = -0.76$ [−1.47, −0.01], $d = -0.32$. A similar effect size was observed for the financial stimuli, though not statistically significant—truthful statements ($M = 5.40$, 95% CI [5.04, 5.76]) were rated as less truthful than deceptive statements ($M = 5.91$, 95% CI [5.42, 6.36]), $M_{diff} = -0.51$, 95% CI [−1.10, 0.09], $d = -0.25$.

Accuracy by judgment type

We further explored judgment accuracy by determining whether participants' accuracy depended on their judgment—labeling a statement as a truth or lie. While overall truth (49.7%) and lie (51.6%) judgments were similarly accurate, $M_{diff} = -0.018$, 95% CI [−0.092, 0.054], two of the stimulus sets had notable patterns. On the Vrij and Mann [2001] videos, truth judgments were incredibly inaccurate (23.9%)—significantly worse than chance ($M = 0.239$, 95% CI [0.134, 0.328]) and significantly worse than the accuracy for lie judgments (61.3%), $M_{diff} = -0.374$, 95% CI [−0.505, −0.236]. The opposite pattern was observed with the financial statements—when participants judged a statement to be true, they were significantly more accurate (60.7%) than if they had merely guessed ($M = 0.607$, 95% CI [0.504, 0.684]) and more accurate than when they thought the statement was a lie (38.2%), $M_{diff} = 0.225$, 95% CI [0.078, 0.362].

Truth bias

While participants were slightly more correct when viewing a video that contained a true statement (54.1%) than a false one (47.2%), the difference was not statistically

significant, $M_{diff} = 0.069$, 95% CI [−.004, 0.140]. This lack of difference is reflective of participants' overall judgment bias—there was a slight but nonsignificant truth bias, with 46.6% of all judgments being lies, $M = 0.466$, 95% CI [0.430, 0.501]. Three sets of stimuli showed differences in accuracy between the truthful and false instances, and thus a significant lie or truth bias. For the Vrij and Mann [2001] videos, participants did much worse than guessing when viewing a truthful video (25.0%) than a video containing a lie (59.8%), $M_{diff} = -.348$, 95% CI [−.479, −.203]; this was also apparent by the significant lie bias—more than two thirds (64.9%) of all decisions were lies, $M = 0.649$, 95% CI [0.580, 0.716]. The Sorochinski et al. [2014] videos had a low lie judgment rate (37.9%), and thus higher accuracies when viewing truthful (59.6%) than deceptive (35.2%) statements, $M_{diff} = 0.243$, 95% CI [0.100, 0.384]. For the financial statements, there was a significant truth bias—only 39.4% of statements were judged to be lies, $M = 0.394$, 95% CI [0.326, 0.462]. This truth bias was apparent in the better-than-guessing accuracy rates when judging a truthful video, $M = 0.602$, 95% CI [0.509, 0.684], and the worse-than-guessing performance on deceptive videos, $M = 0.387$, 95% CI [0.280, 0.500].

Moderators of accuracy

Age was not significantly related to judgment accuracy overall, $r = 0.08$, 95% CI [−0.08, 0.24]. Looking at finance judgments specifically, those who were correct in their evaluation of the quarterly earnings call (M age = 38.6 years old) were not significantly older or younger than were those who were incorrect (M age = 39.6 years old), $M_{diff} = 0.95$, 95% CI [−1.85, 3.84].

If deception detection is valuable skill for financial professionals, and if it is a skill that can be improved by working in the industry, one could expect to find increasing lie detection abilities with increasing on-the-job experience. Few to no participants had either 0–2 years of experience or 30+ years of experience—thus, we collapsed the experience categories into three tiers: low experience (0–10 years), moderate experience (10–20 years), and high experience (20+ years). While overall lie detection accuracy did improve slightly while climbing experience tiers, from low-experience ($M = 0.438$, 95% CI [0.361, 0.521]), to moderate experience ($M = 0.510$, 95% CI [0.432, 0.575]), to high experience ($M = 0.516$, 95% CI [0.408, 0.608]), the differences were not statistically significant, $F(2,152) = 1.07$, $p = 0.347$.

The pattern for judgments on the financial set was similar; logistic regression found that, with the low experience group (52.6% accuracy) as the reference condition, moderately experienced participants were not significantly more accurate (53.7%), odds ratio (OR) = 1.05, 95% CI [0.51, 2.13], $p = 0.903$, nor were the participants in the high experience group (56.7%), OR = 1.18, 95% CI [0.48, 2.90], $p = 0.720$.

Metacognitive realism: The accuracy-confidence relation

We investigated the relationship between lie-detection accuracy and confidence by calculating the percentage of correct judgments at each of the confidence levels available (50% to 100%, in increments of 5%). A correctly calibrated set of participants should self-report confidence levels that align to their average accuracy (e.g., when 80% confident, participants would be expected to be correct 80% of the time). The dashed gray diagonal line in the following calibration curve figures represents accurate calibration. Means above the line reflect a point where participants were more accurate than they were confident, meaning that they displayed underconfidence; points below the line represent overconfidence.

As shown in Figure 3, accuracy did not increase with confidence—it stayed remarkably stable. In fact, at every level of confidence save one, the confidence interval encompassed the chance level (50%)—indicating that regardless of how confident participants were, their accuracy was, overall, poor. The one confidence point that did significantly differ from chance—the 90% confidence level—was significantly worse than guessing (29.8% accuracy), indicating tremendous overconfidence, $M = 0.298$ 95% CI [0.149, 0.386].

It is possible that financial professionals' confidence-accuracy calibration is domain specific—that is, their judgment confidence will be similar to their accuracy, but only for the judgments made in the financial set of stimuli. To test this, we generated calibration curves for each of the four stimuli sets (see Figure 4). Due to a relatively low number of judgments made at high confidence levels (see Figure 5), we combined the 90%, 95%, and

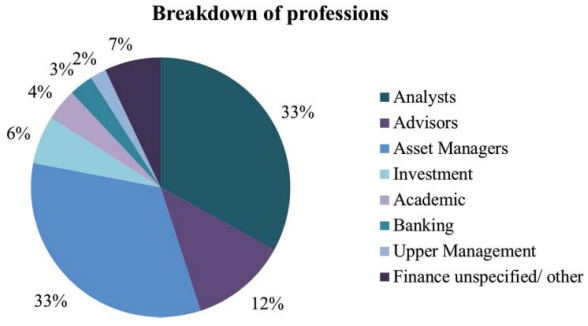

Figure 3. Participants' job categories.

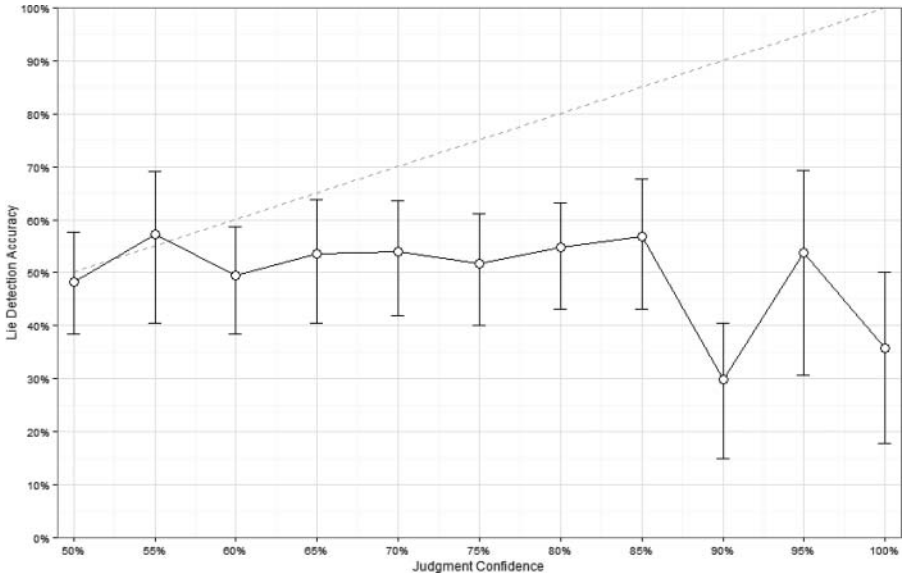

Figure 4. Calibration curve for all participants.

100% confidence judgments together, and grouped the other confidence judgments in chunks of 10%. As shown in Figure 4, performance was similarly poor regardless of the stimuli type. At each level of confidence, our financial professional participants failed to detect financial lies at rates better than chance (50%): at 50% confidence ($M = 0.603$, 95% CI [0.463, 0.721]), at 60% confidence ($M = 0.500$, 95% CI [0.333, 0.619]), at 70% confidence ($M = 0.562$, 95% CI [0.396, 0.666]), and at 80% confidence ($M = 0.438$, 95% CI [0.250, 0.594]) accuracy did not significantly differ from mere guessing. Participants performed especially poorly when they were highly confident. Only one third of financial judgments at confidences of 90% and higher ($n = 18$) were correct ($M = 0.333$, 95% CI [0.111, 0.500]), significantly worse than they would have achieved by flipping a coin.

Discussion

In this study, we set out to examine investment professionals' decision making regarding deception. More specifically, we were interested in these professionals' ability to distinguish between truthful and deceptive statements. We were also interested in the extent to which investment professionals display a judgment bias. As we will discuss further, data on judgment biases are informative

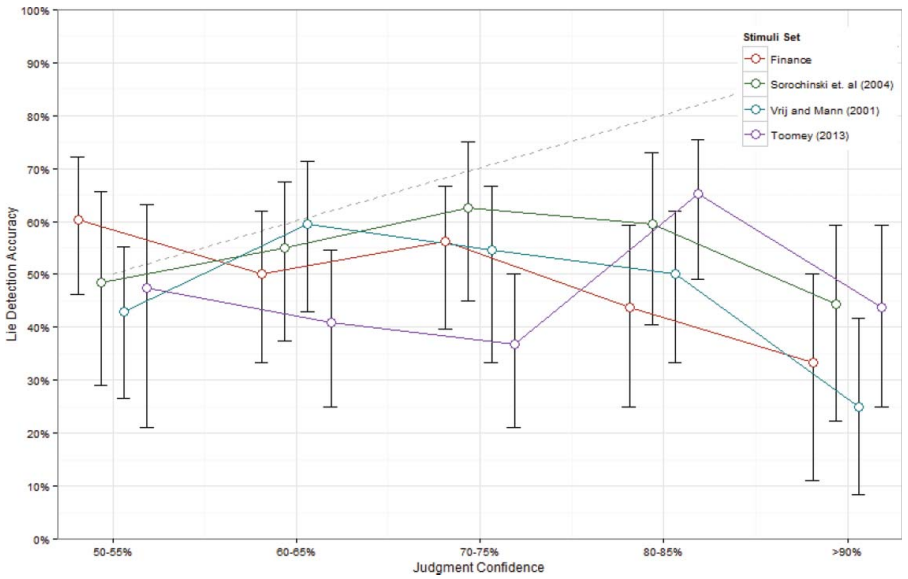

Figure 5. Calibration curves for each of the four stimuli.

as they provide information of what type of error a given group is likely to fall prey to when attempting to detect deception. Finally, we examined the metacognitive realism of investment professionals.

In order to subject investment professionals' judgments to an experimental test, we exposed them to truthful and deceptive statements derived from a variety of populations and settings. First, we used relatively lengthy statements provided during a mock investigative interview, in which people lied or told the truth about their involvement in a transgression. Second, we used statements given by convicted felons during the course of a laboratory study. Third, we included a sample of high-stakes lies and truth told by people during press conferences with the public. Fourth, we used truthful and fraudulent statement given during conference calls. The latter sample is of particular importance because it represents a setting in which investment professionals are often required to assess the reliability of the information provided.

For all statements, we had ground truth (unambiguous knowledge about the veracity of the statements); hence, we could calculate accuracy in participants' judgments. Further, our methodological approach offers two advantages in terms of external validity (i.e., the ability to generalize the results from this study to other contexts and settings). First, given the variation in the types of statements we exposed participants to, we can be reasonably confident that our results translate to other types of lies. Second, because we exposed participants to statements from their own professional domain in the form of fraudulent conference calls, we are able to draw some conclusions about their performance in their work life.

Major findings on lie detection accuracy

The most important finding emerging from this study is that financial professionals are poor at distinguishing between true and false statements. We found an average accuracy rate of 49.4% across the different types of statements we exposed them to. Our participants were operating on the very same accuracy rate one would obtain from simply guessing or flipping a coin. Yet, from the perspective of the general deception detection literature, this poor performance is not shocking (see Bond and DePaulo [2006]).

How can it be that these professionals displayed a chance-level performance when faced with this important task? Here, we focus on three plausible explanations for our results. First, the literature shows that people lie on a regular basis (DePaulo et al. [1996], DePaulo and Kashy [1998]). It thus seems like deception is part of a larger set of cognitive skills that help people function in social life. In the so-called self-presentational perspective on deception (DePaulo et al. [2003], see also DePaulo [1992]), lying is not an extraordinary activity. According to this perspective, we constantly edit the way we come across in order to reach desirable goals. For example, in a job interview, we dress, act, and speak in ways that are meant to communicate our competence and fit for the job. In other settings (e.g., on a first date, in family gatherings, socializing with friends, etc.), we might present and highlight quite different aspects of ourselves. In sum, research suggests that we can expect people to be quite skilled at providing deceptive statements, primarily due to practice with similar tasks.

Second, the literature on cues to deception suggests that lie catchers face a very difficult task (DePaulo et al. [2003]). In statistical terms, the average effect size for the association between behavioral indicators and deception is very small—in a subset of their analyses, DePaulo et al. found an effect size of $d = 0.10$. The traditional way to interpret effect sizes is to categorize those that exceed $d = 0.20$ as small effect, with those over 0.50 as medium, and 0.80 as large (Cohen [1988]). An example of a small effect size is the average difference in height between 15- and 16-year-old girls. This is a real difference, but it is small and in all likelihood very difficult to detect with the naked eye. In some ways then, attempting to detect lies appears similar to the task of trying to diagnose a physical disease which has exceedingly weak observable symptoms: It should come as no surprise that people faced with this task fall prey to error.

Third, to compound the problem, people harbor a variety of misconceptions about the nature of deceptive behavior (e.g., Global Deception Research Team [2006]). A recent survey demonstrated that investment professionals share these misconceptions. In light of the finding that investment professionals subscribe to false stereotypes about deceptive behavior, poor accuracy in lie judgments is not surprising.

Moderators of lie detection accuracy

We tested investment professionals' ability to detect lies derived from a variety of domains, including the financial one. Common sense suggests that participants ought to have performed better when judging statements of a form with which they were presumably familiar; that is the sample of lies and truth told in the financial realm. It also seems reasonable to think that the stakes of the situation in which the lie is told may play a role. More specifically, when liars face serious consequences if they fail to convince, it is conceivable that their lies are more detectable. In this study, we found that while accuracy was marginally better for statements that involved financial

fraud –and high stakes- than statements from other context, the difference was nonsignificant from a statistical point of view, and trivial from a practical point of view. Our finding thus mirrors those obtained in the larger literature: Professional expertise and familiarity does not seem to produce domain-specific lie detection skills; and lies are in fact not more easily detected when they are told under high-stakes, highly emotional conditions (Hartwig and Bond [2014]).

We did not find that age or years of experience working in the financial industry moderated accuracy. In order to understand how it might be plausible that lie detection skills do not develop over time, it is useful to invoke the basic literature on learning and decision making. Hogarth [2001] introduced a framework to understand the conditions under which skills are likely to improve as a function of experience. He suggested that two elements need to be present in order for people to learn the right lessons from experience: First, the consequences of misjudgments must be severe (in order to motivate people to learn from their mistakes). Second, there needs to be consistent, timely and reliable outcome feedback on judgments. Consider the structure of the environment in which investment professionals operate: The consequences of failing to make accurate judgments of deception may indeed be severe. However, because there is no process through which investment professionals receive immediate and clear information about their judgments of deception, the kind of feedback that may be necessary for learning is often absent. Hogarth [2001] labeled those domains in which the consequences of misjudgments are severe, but feedback is lacking as wicked learning structures. In such an environment, one cannot expect people to learn from their experience, since the mechanisms that provide people with the necessary feedback to adjust their decision-making rules are absent.

Judgment bias

As described in the introduction, people generally display a truth bias. Judgment biases have an impact on lie and truth detection accuracy (Levine et al. [2006]). Researchers have identified a so-called veracity effect, which means that people are more accurate at identifying statements that are actually truthful compared to those that are actually deceptive (Levine, Park, and McCornack [1999]).

We do not know the base rate of truths and lies in any setting, including the financial realm. In this study, we employed the standard procedure of a base rate of 50/50 truths and lies, and found that there was a nonsignificant tendency toward a truth bias across all the stimulus materials. However, it is interesting to note the patterns of judgment bias broken down by stimulus type. In particular, we found that for the lies derived from the financial statements displayed a marked truth bias—more than 60% of all statements were deemed to be truthful. This led to a veracity effect for the financial statements, where accuracy was higher than chance for statements that were actually true, but significantly worse than chance for deceptive statements. This is an unusual finding, as most previous research on presumed lie experts shows a lie bias. There are several possible interpretations of this finding. First, it may be that investment professionals are not particularly wary of the possibility of being duped in their work environment, and that they therefore operate on a default assumption of truth when they come across messages from their professional life (Gilbert [1991]). A second possibility is that the fraudulent statements given during the conference calls were particularly believable, and that other groups too would display a truth bias for these statements. Without further research, we cannot confidently conclude which of these explanations are more likely to be true.

In sum, while the accuracy rates we observed were in line with previous research on deception detection, our analyses of judgment bias raise several important questions worthy of further empirical scrutiny. In particular, the pronounced truth bias for the financial statements, and the low accuracy rates for detecting lies that followed from the truth bias are puzzling. As we discuss subsequently, these findings may have important practical implications.

Metacognitive realism

In order to further examine the decision making of investment professionals, we analyzed their metacognitive realism. Our analyses reveal exceedingly poor metacognitive realism: Participants' accuracy rates were largely similar across the different confidence intervals, suggesting that they had very poor insight into their own decision making. Moreover, there was a marked tendency for overconfidence, meaning that participants expressed far higher confidence levels than their judgment accuracy warranted. The picture painted by our breakdown of realism by stimulus material was equally bleak: For the financial statements, only a third of the judgments made with extremely high confidence were correct. This means that for the domain in which our participants ought to have special expertise, their metacognitive calibration was the opposite of what one would

expect from rational judgments—when they made judgments with a very high level of confidence, they were significantly less accurate than chance.

Our findings are consistent with the literature on metacognitive calibration in several ways: First, metacognitive realism tends to be poor across a variety of domains (Lichtenstein and Fischhoff [1977]). Second, people usually display overconfidence (Overschelde [2008]). Third, the finding that highly confident participants were less rather than more accurate is reminiscent of the so-called Dunning-Kruger effect (Dunning [2011], Kruger and Dunning [1999]). This is a pattern in which people with less accurate decision making are more overconfident than are those who are more skilled.

Practical implications

While the finding about investment professionals' poor lie detection accuracy is not shocking based on our previous knowledge of human lie judgments, it is still alarming from a practical point of view. Overall, we can expect them to make both false positives (e.g., mistakenly believing that an honest statement is a lie) and false negatives (e.g., failing to detect a deceptive statement, instead judging it to be truthful). Our results from the financial stimuli suggest that false negatives are more likely. Yet, as discussed previously the history of business is replete with fraudulent activity, and most such activity escapes even the notice of auditors. In these situations typical forms of analysis—such as financial statement analysis, financial ratios, site visits to businesses, and channel checks—are unlikely to uncover fraudulent activity. This puts particular pressure on financial analysts' ability to judge the veracity of statements made by executives about the businesses they manage. Our study has identified one mechanism through which fraud can be perpetrated and remain undetected—through the failure in judgments by investment professionals.

Limitations and future directions

While we believe that this study offers novel and compelling data of significant practical importance, there are some limitations that should be noted. First, while we used a broad sample of lies and truth derived from a variety of settings, it could be argued that we placed the sample of lie catchers in a difficult situation: They did not have much background information about the statements they encountered, and they were confined to observing and listening to the statements without the opportunity to pose question. This is the typical laboratory paradigm in the deception literature (Hartwig [2011]). It is possible that participants would have performed better if they had the opportunity to plan and pose questions to the people whose veracity they had to judge. However, given past research showing that even professional interviewers do not fare better when allowed to interact with the subject (Hartwig et al. [2004]), there is not much reason to expect that investment professionals would perform better if they had assumed a more active role. Still, it is a question for future research whether the accuracy rates obtained here would hold true across other judgment situations.

Further, this type of paradigm ignores the possibility of repeated interactions between would be liars and judges. This type of sequential interaction would allow for judges to form impressions of trustworthiness about the speakers (e.g., Sobel [1985]). This is especially important when considering financial decision making as research in the past has shown that judgments of credibility can be influenced by a sender's repeated behavior (e.g., Lunawat [2013], Benabou and Laroque [1992]). However, these studies take into account impressions of the speaker and not the actual veracity of the statement. Future researchers should incorporate judges' prior interactions with a sender as a factor of consideration when testing deception detection accuracy.

Some of the findings in our study were surprising, in particular the pattern of judgment biases observed across the different types of lies. We have speculative explanations for these patterns, but we do not have the data necessary to test any of these. We hope future researchers will (a) examine whether the findings can be replicated and, if so, (b) provide firmer explanations of the psychological mechanisms behind these judgment biases.

Conclusions

In this study, we demonstrated that investment professionals operate on chance-level accuracy when faced with the task of detecting deception. This finding held true across a variety of real-life and laboratory statements, and across low- and high-stakes settings. Our findings on investment professionals' judgment biases suggest that they are particularly credulous when it comes to judging statements from the financial domain, and this credulity leads to a very low accuracy rate for detecting statements that were actually deceptive. Our analyses of the confidence judgments made in relation to judgments of deception demonstrate poor metacognitive calibration. This shows that investment professionals have limited insight into their own decision making, which makes it hard or even impossible for them to know when they should act on their judgments of deception. This is even more concerning because research does show that they do rely on these judgments to make

financial decisions (e.g., Lunawat [2013], Benabou and Laroque [1992]).

The picture emerging from our analyses is not an optimistic one: The combination of poor accuracy, a tendency for credulity for statements in the financial realm, and poor metacognitive realism is highly problematic, and suggests that we can expect a very high error rate in investment professionals' judgments of deception. We urge future research to further explore the decision making made by investment professionals, and in particular to consider empirically based methods to improve accuracy.

Acknowledgments

The authors are thankful to the CFA Institute for their assistance with data collection. We are particularly indebted to Prashant Goswami, Melissa Looney, and Lisa Schuetz. Also, we are thankful to Bloomberg, L.P., especially Andy Demeter and Jim Ziegler for helping us generate stimulus materials.

Notes

1. For more information on how guilt was established, see Vrij and Mann [2001].

References

Anderson, N. H. "Likableness Ratings of 555 Personality-Trait Words." *Journal of Personality and Social Psychology*, 9, (1968), pp. 272–279.

Backbier, E., J. Hoogstraten, and K. Meerum Terwogt-Kouwenhoven. "Situational Determinants of the Acceptability of Telling Lies." *Journal of Applied Social Psychology*, 27, (1997), pp. 1048–1062.

Bond, C. F. Jr. and B. M. DePaulo. "Accuracy of Deception Judgments." *Personality and Social Psychology Review*, 10, (2006), pp. 214–234.

Benabou, R. and G. Laroque. "Using Privileged Information to Manipulate Markets Insiders, Gurus, and Credibility." *The Quarterly Journal of Economics*, 107, (1992), pp. 921–958.

Carver, R. P. "The Case Against Statistical Significance Testing." *Harvard Educational Review*, 48, (1978), pp. 378–399.

Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edition. Hillsdale: Erlbaum, 1988.

DePaulo, B. M. "Nonverbal Behavior and Self-presentation." *Psychological Bulletin*, 111, (1992), pp. 203–243.

DePaulo, B. M., K. Charlton, H. Cooper, J. J. Lindsay, and L. Muhlenbruck. "The Accuracy-confidence Correlation in the Detection of Deception." *Personality and Social Psychology Review*, 1, (1997), pp. 346–357.

DePaulo, B. M., J. A. Epstein, and M. M. Wyer. "Sex Differences in Lying: How Women and Men Deal with the Dilemma of Deceit." In M. Lewis and C. Saarni, eds., *Lying and Deception in Everyday Life* (pp. 126–147). New York: Guilford Press, 1993.

DePaulo, B. M. and D. A. Kashy. "Everyday Lies in Close and Casual Relationships." *Journal of Personality and Social Psychology*, 74, (1998), pp. 63–79.

DePaulo, B. M., D. A. Kashy, S. E. Kirkendol, M. M. Wyer, and J. A. Epstein. "Lying in Everyday Life." *Journal of Personality and Social Psychology*, 70, (1996), pp. 79–995.

DePaulo, B. M., J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper. "Cues To Deception." *Psychological Bulletin*, 129, (2003), pp. 74–118.

Dunning, D. "The Dunning-Kruger Effect: On Being Ignorant of One's Own Ignorance." In M. P. Zanna and J. M. Olson, eds., *Advances in Experimental Social Psychology, Vol. 44* (pp. 247–296). San Diego: Academic Press, 2011.

Garrido, E., J. Masip, and C. Herrero. "Police Officers' Credibility Judgments: Accuracy and Estimated Ability." *International Journal of Psychology*, 39, (2004), pp. 254–275.

Gilbert, D. T. "How Mental Systems Believe." *American Psychologist*, 46, (1991), pp. 107–119.

Global Deception Research Team. "A World of Lies." *Journal of Cross-Cultural Psychology*, 37, (2006), pp. 60–74.

Hartwig, M. "Methods in Deception Research." In B. Rosenfeld and S. Penrod, eds., *Research Methods in Forensic Psychology* (pp. 136–155). Chichester: Wiley, 2011.

Hartwig, M. and C. F. Bond Jr. "Why Do Lie-Catchers Fail? A Lens Model Meta-analysis of Human Lie Judgments." *Psychological Bulletin*, 137, (2011), pp. 643–659.

Hartwig, M. and C. F. Bond Jr. "Lie Detection from Multiple Cues: A Meta-analysis." *Applied Cognitive Psychology*, 28, (2014), pp. 661–676.

Hartwig, M. and P. A. Granhag. "Exploring the Nature and Origin of Beliefs About Deception: Implicit and Explicit Knowledge Among Lay People and Presumed Experts." In P. A. Granhag, A. Vrij, and B. Vershuere, eds., *Deception Detection: Current Challenges and New Directions* (pp. 125–153). Chichester: Wiley, 2015.

Hartwig, M., P. A. Granhag, L. A. Strömwall, and A. Vrij. "Police Officers' Lie Detection Accuracy: Interrogating Freely Versus Observing Video." *Police Quarterly*, 7, (2004), pp. 429–456.

Hartwig, M., J. A. Voss, and D. B. Wallace. "Detecting Lies in the Financial Industry: A Survey of Investment Professionals' Beliefs." *Journal of Behavioral Finance*, 16, (2015), pp. 173–182.

Hogarth, R. M. *Educating Intuition*. Chicago: University of Chicago Press, 2001.

Kassin, S. M., R. A. Leo, C. A. Meissner, K. D. Richman, L. H. Colwell, A.-M. Leach, and D. La Fon. "Police Interviewing and Interrogation: A Self-report Survey of Police Practices and Beliefs." *Law and Human Behavior*, 31, (2007), pp. 381–400.

Kruger, J., and D. Dunning. "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-assessments." *Journal of Personality and Social Psychology*, 77, (1999), pp. 1121–1134.

Levine, T. R., R. K. Kim, H. S. Park, and M. Hughes. "Deception Detection Accuracy is a Predictable Linear Function of Message Veracity Base-rate: A Formal Test of Park and Levine's Probability Model." *Communication Monographs*, 73, (2006), pp. 243–260.

Levine, T. R., H. S. Park, and S. A. McCornack. "Accuracy in Detecting Truths and Lies: Documenting the 'Veracity Effect.'" *Communication Monographs*, 66, (1999), pp. 125–144.

Lichtenstein, S., and B. Fischhoff. "Do Those Who Know More Also Know More About How Much They Know?" *Organizational Behavior and Human Performance*, 20, (1977), pp. 159–183.
Lunawat, R. "An Experimental Investigation of Reputation Effects of Disclosure in an Investment/trust Game." *Journal of Economic Behavior and Organization*, 94, (2013), pp. 130–144.
Meissner, C. A., and S. M. Kassin. "He's guilty!": Investigator Bias in Judgments of Truth and Deception. *Law and Human Behavior*, 26, (2002), pp. 469–480.
Overschelde, J. P. V. "Metacognition: Knowing About Knowing." In J. Dunlosky and R. A. Bjork, eds., *Handbook of Metamemory and Memory* (pp. 47–71). New York: Psychology Press, 2008.
Schwarz, N. "Metacognition." In M. Mikulincer, P. R. Shaver, E. Borgida and J. A. Bargh, eds., *APA Handbook of Personality and Social Psychology, Volume 1: Attitudes and Social Cognition* (pp. 203–229). Washington, DC: American Psychological Association, 2015.
Serota, K. B., and T. R. Levine. "A Few Prolific Liars: Variation in the Prevalence of Lying." *Journal of Language and Social Psychology*, 34, (2015), pp. 138–157.
Sobel, J. "A Theory of Credibility." *The Review of Economic Studies*, 52, (1985), pp. 557–573.
Sorochinski, M., M. Hartwig, J. Osborne, E. Wilkins, J. Marsh, D. Kazakov, and P. A. Granhag. "Interviewing to Detect Deception: When to Disclose the Evidence?" *Journal of Police and Criminal Psychology*, 29, (2014), pp. 87–94.
Strömwall, L. A., P. A. Granhag, and M. Hartwig. "Practitioners' Beliefs about Deception." In P. A. Granhag, and L. A. Strömwall, eds., *The Detection of Deception in Forensic Contexts* (pp. 229–250). Cambridge: Cambridge University Press, 2004.
Toomey, J. A. "Understanding the Role of Psychopathic Characteristics in Deceptive Behavior. Unpublished study, 2013.
U.S. Equal Employment Opportunity Commission. "Diversity in the Finance Industry" (2006). Retrieved from http://www.eeoc.gov/eeoc/statistics/reports/finance/index.html
Vrij, A. *Detecting Lies and Deceit: Pitfalls and Opportunities*, 2nd Edition. New York: John Wiley and Sons Ltd., 2008.
Vrij, A., and S. Mann. "Who Killed my Relative? Police Officers' Ability to Detect Real-life High-stake Lies." *Psychology, Crime and Law*, 7, (2001), pp. 119–132.

Appendix

Descriptions of the conference calls

Indy Mac Bancorp. These recordings took place during Indy Mac's first quarter of 2008 conference call. During the deceptive call, IndyMac Bancorp's corporate executive presented quarterly earnings, claimed to be well capitalized and predicted they would continue to do so throughout 2008. In one of the truthful statements, Indy Mac's corporate executive claimed that their year to year losses were coming down and that these losses were manageable. In the second truthful statement, the corporate executive claimed that their net losses were down and that production had improved. He also predicted that production would break even in the second quarter and be modestly profitable in the second half of 2008.

Hansen Medical. These statements were taken from Hansen Medical's third quarter conference calls, announcing earnings for 2008. In their deceptive statement, the corporate executive for Hansen Medical reported record increase in quarterly revenue He also detailed what lead to this increased income and reported that this was their highest average sales price to date. In their truthful statement, the corporate executive reported an increase in research and development expenses and projected an additional increase in these expenses over the year.

Office Depot. The statements for this company were taken from the earnings conference call for the fourth quarter of 2004. In the deceptive statement, the corporate executive stated that they had achieved a growth in earnings per share in 2006 and over the past two years. She also claimed that Office Depot could deliver top line growth in 2007 and predicted some margin expansion each year, annual and cost leverage, and that their business model should contribute solid EPS growth. In the first truthful statement, Office Depot's corporate executive stated that their North American business solutions division sales increased revenue and reflected growth compared to the previous year. She then stated that although contract sales growth was somewhat depressed they expected to see a continuation of sales compression in 2007. In the second truthful statement, the corporate executive stated that the North American business solutions division had a lower operating profit for the fourth quarter of 2006 compared to the same period the prior year. She provided details of the expenses that raised operating costs in the fourth quarter and were the primary contributor of margin erosions but added that they were expected to moderate over the next few quarters due to several factors that she detailed.