Text-Based Unethical Behavior Forecasting:

The Hidden Information Distribution and Evaluation (HIDE) Model

Yeonjeong Kim

Organizational Behavior & Theory

Tepper School of Business

Carnegie Mellon University

Dissertation Committee:

Professor Taya R. Cohen (Chair)

Professor Laurie R. Weingart

Professor Denise M. Rousseau

Professor Rebecca L. Schaumberg

Text-Based Unethical Behavior Forecasting:

The Hidden Information Distribution and Evaluation (HIDE) Model

Abstract

One of the biggest problems facing organizations is unethical employee behavior such as cheating and stealing. One way to effectively mitigate unethical work behavior is to identify unethical individuals during the selection process. However, it is currently unknown whether, or how, we can detect peoples' tendency to behave unethically when we do not know the person well. This research is designed to remedy this dearth in our understanding of unethical behavior predictions in settings where people need to make prompt judgments based on the limited information they obtain from strangers.

In Chapter 1 of this dissertation I develop a new theoretical framework, the hidden information distribution and evaluation (HIDE) model. This model predicts that judges, who do not know the target individuals of evaluation, can detect aspects of unethical behavior tendencies that targets incorrectly know (misconstrue) and/or are unaware of themselves. Using this model, I developed a novel tool to predict the unethical behavior of people from their spontaneous written responses to specially designed questions.

In Chapter 2, I conducted laboratory experiments and a field survey to investigate the wisdom of crowds in forecasting unethical behavior from written interview responses of targets. I show that groups of naïve judges can predict the unethical behavior of targets by evaluating their moral character using this text-based interview method.

In Chapter 3, I investigate what aspects of moral character are revealed in each interview question with an aim to further increase the predictive power of unethical behavior using the

text-based interview method. To increase the predictive validity, I found that certain evaluation dimensions should be matched to particular interview questions because each question revealed different aspects of moral character. Across three studies, the judges' evaluations of more specifically defined moral character traits (i.e., Conscientiousness, guilt proneness) had better convergent validity and stronger predictive powers than the judges' evaluations of moral character as a whole. Additionally, I found that the judges' evaluations of Honesty-Humility were not as predictive as other dimensions.

In Chapter 4, I investigated the predictive validity of the judges' evaluations with varying levels of the targets' impression management motivation when answering the interview questions. The relative predictive powers of the judges' ratings, compared to self-reports, increased as the targets' impression management motivation increased. When high levels of impression management were employed, only the reports by the judges were predictive of the unethical behavior by targets. In Chapter 4, I also investigated how judges form an impression of the Honesty-Humility of targets. I found that the judges' evaluations of the four elements of the Honesty-Humility factor (i.e., sincerity, fairness, greed-avoidance, and modesty) do not form one factor as the greed element was somewhat positively perceived in judging others and was positively correlated with Conscientiousness evaluations.

In Chapter 5 I conducted text analyses to explore how human judges utilize linguistic cues in written responses to form an impression of moral character and how linguistic cues predict the unethical behavior of targets. The goal of this final chapter is to detect the linguistic cues that human judges failed to correctly detect or utilize. I introduced the future direction of this research program using exploratory text analyses.

Key words: unethical behavior; interviews; text-analysis; person perception.

CHAPTER I

The Hidden Information Distribution and Evaluation (HIDE) Model

One of the biggest problems facing organizations is employee unethical behavior, such as cheating and stealing (Dalal, 2009, Kim, Cohen, & Panter. 2016). One way to effectively mitigate unethical work behavior is to identify unethical individuals during the selection process (Kim & Cohen 2015; Kim et al. 2016). However, it is currently unknown whether, or how, we can detect peoples' tendency to behave unethically if we do not know the person well. The goal of this dissertation is to answer the question of whether, and how to, people can evaluate strangers' tendency to behave unethically in situations where those judges (e.g., interviewers) need to make prompt evaluations based on a limited set of information about those strangers (e.g., job candidates).

Individuals' dispositions toward behaving unethically are studied in the literature of moral character. Specifically, recent psychological research has approached the study of moral character from a personality perspective, which posits that moral character is composed of characteristic patterns of thought, emotion, and behavior that are associated with morality and ethics (Cohen & Morse, 2014; Cohen, Panter, Turan, Morse, & Kim, 2015; Fleeson, Furr, Jayawickreme, Meindl, & Helzer, 2014; Kim & Cohen, 2015; Lee & Ashton, 2012; Peterson & Seligman, 2004). This work defines *personality* as "an individual's characteristic patterns of thought, emotion, and behavior, together with the psychological mechanisms— hidden or not— behind those patterns" (Fast & Funder, 2010, p. 669). *Personality traits* are unobservable psychological constructs that encapsulate patterns of thought, emotion and behavior into coherent units, and thus facilitate understanding of how individuals differ from one another (Fast & Funder, 2010). The terms "morality" and "ethics", refer to standards of right and wrong

conduct that provide guidance on what we should and should not do. In particular, prior researchers argue that helpful acts are hallmarks of ethical/moral behavior while harmful acts are hallmarks of unethical behavior (Cohen et al., 2014).

In this dissertation, I define moral character more conservatively by restricting its scope. Specifically, I define moral character as an umbrella term referring to a subset of personality traits that predict individuals' *unethical behaviors* consistently across diverse situations. That is, I define moral character using unethical behavior as the sole criterion rather than including both ethical and unethical behavior.

The reason I define moral character using solely unethical behavior is because the concept of ethical or unethical behavior includes the autonomy (Hogan, 1973), evaluative value (Schwartz et al., 2012), or motivational elements of the behavior (Cohen et l., 2014; Cohen & Morse, 2014). This means that whether a behavior is right or ethical cannot be evaluated without considering the basic reason why the actor engages in such conduct. When the motivation is purely self-benefiting rather than other-benefiting, the act is not considered ethical regardless of whether the behavior seems helpful to others on a surface level. For example, right conduct (e.g., helping others) can be sourced back to one's self-benefiting motivations such as cultivating social networks or building positive reputations, in addition to stemming from social norms. In contrast, wrong conduct (e.g., harming others) is less likely to be interpreted as having other-benefiting motivations. Therefore, right conduct (i.e., ethical behavior) is more interpretative and ambiguous in its motivations than unethical behavior. Considering that, in everyday life, the scope of right conduct is more difficult to define clearly than unethical behavior, I focus on unethical behavior in defining moral character.

Understanding individual differences in moral character allows us to predict and possibly prevent unethical behaviors that harm people, organizations, and society (Kim & Cohen, 2015). Indeed, measures that capture information relevant to moral character reliably predict observable unethical behaviors in anonymous research settings. For example, self-reports of Honesty-Humility—one of the "Big Six" factors from the HEXACO model of personality structure, which encompasses sincerity, fairness, modesty, and greed-avoidance—predicts not only self-reported delinquency and unethical decision but also observable dishonesty, such as in behavioral economics games (Hilbig & Zettler, 2015), and coworker-reported workplace deviance (Cohen, Panter, Turan, Morse, & Kim, 2013). Other-reports of Honesty-Humility also predict self-reported delinquency and unethical decisions as well as coworker-reported workplace delinquency (Cohen et al., 2013). Likewise, self-reported guilt proneness—an individual difference indicative of whether a person would feel guilty about committing transgressions even if no one were to find out—also predicts self-reported and observable unethical behaviors (Cohen, Kim, Jordan, & Panter, 2016; Cohen, Wolf, Panter, & Insko, 2011), and both self- and other-reports of guilt proneness predict self-reports and coworker-reports of workplace deviance (Cohen et al., 2013). Even more striking is the observation that guilt proneness measured with self-reports in children aged 10 to 12 correlates negatively with illegal behavior during young adulthood and with involvement in the criminal justice system through ages 18 to 21, providing powerful evidence of the importance of this moral character trait for predicting consequential harmful behaviors (Stuewig et al., 2015).

Although previous research has clearly shown that self-reported moral character traits, as well as assessments made by well-acquainted others, predict unethical behaviors, we currently do not know whether we can accurately evaluate *strangers' moral character*, nor do we know *how*

to elicit relevant information from strangers. The biggest challenge in assessing a person's moral character is that it is an extremely evaluative trait (i.e., high in social desirability)—if not the most evaluative trait. This is because moral character plays a central role in shaping how we view ourselves (Fernandez-Duque & Schwartz, 2016) as well as how others view us (Goodwin, Piazza, & Rozin, 2014; Goodwin, 2015).

When a trait is evaluative, self-perceptions are often distorted because of ego-protection motivation (Asendorpf & Ostendorf, 1998; Vazire, 2010). Therefore, whether consciously or unconsciously, individuals are likely to fall prey to self-deception and impression management. This means that when judging strangers' moral character, we need a way to reveal aspects of targets' moral character beyond what those targets report themselves, especially aspects that they are unaware of themselves or able to control.

As a first step toward answering the question of how we can make valid judgments of strangers' moral character, in Chapter 1 of this dissertation, I develop the *hidden information distribution and evaluation* (HIDE) model, which posits that self-reports and other-reports each capture unique insights about the targets of judgment because certain kinds of information are hidden from one party and detectable only by the other. Applying the HIDE model to moral character judgments, I propose that judges who do not know the targets are able to detect aspects of moral character that target individuals incorrectly know (misconstrue) and/or are unaware of themselves. To elicit information about targets' moral character of strangers, this research develops character interview questions that are designed to reveal targets' moral character through their spontaneous written responses to interview questions. I propose that impromptu thinking and language usage in answering these questions reveal information about moral

character that targets are unaware of themselves and thus less able to control but that judges can use to make valid character judgments that are predicticve of targets' unethical behaviors.

The HIDE model and its implications for moral character judgments have the potential to make groundbreaking theoretical and applied contributions to organizational psychology and related fields. For example, in many interview settings, judges (e.g., potential employers) are limited to evaluating targets' (i.e., job candidates') moral character from small samples of linguistic cues from their responses to interview questions. Yet, we currently do not know how to elicit particularly relevant linguistic cues from targets, nor do we know whether character judgments based on verbal and/or written linguistic cues are diagnostic of unethicality. These are critical issues for organizations considering that interview methods are a centerpiece of employee selection procedures (Huffcutt, Iddekinge, & Roth, 2011) and that moral character judgments can be an important means to identify individuals who might harm (or help) organizations and the people within them. More broadly, this research paves the way toward increased theoretical development in our understanding of what moral character is, how it is revealed in written responses to interview questions, and how to assess it.

The Relative Accuracy of Self- and Other-Perceptions of Personality

The field of personality psychology has largely been built on targets' self-reports (Connelly & Hűlsheger, 2012; Oh, Wang, & Mount, 2011). Indeed, self-reported personality has been shown to predict individuals' own behaviors and life outcomes to a remarkable degree (Funder & Colvin, 1991; Ozer & Benet-Martinez, 2006; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). It is obvious that the self has an advantage in accessing information that might not be observable to others (e.g., affect). However, a wealth of empirical research demonstrates that for certain traits, assessments made by others outperform self-reports in predicting relevant

behaviors and life outcomes, especially when the criterion is provided by a third party (i.e., neither the target nor judges who provide other-reports) or measured objectively (e.g., Asendorpf & Ostendorf, 1998; Connelly & Hűlsheger, 2012; Gosling, John, Kenneth, & Robins, 1998; Kolar, Funder, & Colvin, 1996; Oh, Wang, & Mount, 2010; Vazire, 2010; Vazire & Mehl, 2008). For example, Connelly and Hűlsheger (2012) found that personality measured by other-reports were more predictive of targets' job performance than self-reports.

To understand the relative validity of self- and other-reports of personality, several theoretical frameworks have been proposed. For example, the Johari Window (Luft & Ingham, 1955) partitions personality knowledge into four categories: aspects that the target and others both know (arena), aspects that only the target knows (facade), aspects that only others know (blind spot), and aspects that neither knows (unknown). Building on the Johari Window, Vazire (2010) developed the self-other knowledge asymmetry (SOKA) model, suggesting that targets are more accurate than others in judging traits that are low in observability (e.g., neuroticism) but that others are more accurate than targets when the traits are evaluative (i.e., highly socially desirable; e.g., intellect-related traits). In general, empirical studies in this area support the idea that evaluativeness and observability are important determinants of accuracy of self- and other-perceptions across traits (e.g., Asendorpf & Ostendorf, 1998; Connelly & Ones, 2010; Connelly & Hűlsheger, 2012; Gosling, John, Kenneth, & Robins, 1998; Human & Biesanz, 2011; Kolar, Funder, & Colvin, 1996; Oh, Wang, & Mount, 2010; Vazire, 2010; Vazire & Mehl, 2008).

Funder's (1995; 2012) Realistic Accuracy Model (RAM) model provides the theoretical framework to understand how an accurate personality judgment can happen. The RAM model describes four necessary steps for an accurate personality judgment. First, relevance—the target must provide relevant cues to the trait being judged. Second, availability—the trait-relevant

information must be available to the judge. For example, judges need to have an opportunity to observe targets' behaviors that are associated with the focal trait. Third, detection—the judge must be able to detect available and relevant information about the trait, meaning that the judge must have sufficient ability and motivation to see and understand the information and not ignore it. Finally, utilization—the judge must use the trait-relevant, available, and detected information correctly and not misinterpret it. Each condition of this process influences the extent to which the target's trait is connected to the judge's correct evaluation of that trait. Therefore, validity in personality judgments is likely to be high when the information provided is strong in quantity and quality ("good information"), the focal trait is visible and easily judged ("good trait"), the target is judgeable ("good target"), and the judge is well-calibrated ("good judge") (Funder, 2012). The "good trait" component of the RAM is connected to the SOKA model (Funder, 2012). In the RAM model, trait evaluativeness is detrimental to accurate person perception because self-deception and impression management tactics distort availability and relevance of cues. Trait observability, on the other hand, improves the accuracy of person perception because more visible traits are more available to judges and easier to detect, hence judges' evaluations are more likely to be accurate (Funder, 1995).

The study of accuracy in personality judgments has greatly advanced our understanding of *between-trait differences* in self-other perceptions such that when self- and other-reports are dissimilar, one of rating source is inferred to be more accurate based on the trait's evaluativeness and observabilty (e.g., Asenrdorpf & Ostendorf, 1998; Gosling et al., 1998; Vazire, 2010). However, currently, the field lacks understanding of *the distribution of information within a trait* that is detectable from self- and other-perceptions. In particular, an important but rarely studied theoretical possibility that this dissertation focuses on is that self- and other-reports both provide

partially valid information about targets, but they provide different aspects of information from one another.

The Hidden Information Distribution and Evaluation (HIDE) Model

The HIDE model is presented in Figure 1. At its highest level, the HIDE model separates person perception into two evaluation sources: self and judge. In the HIDE model, "judges" refers to those people who provide other-reports (as opposed to self-reports) of the targets being evaluated. These judges could be the targets' acquaintances or they could be strangers to the targets. The model assumes that there is information about a target that can be judged (correctly or not) by the self, and there is information about the target that can be judged (correctly or not) by others (i.e., judges). For each evaluation source, the information of interest (i.e., evaluation domain) about the target is distributed into three non-overlapping components: 1) valid information (*correctly-identified information*); 2) invalid information (*incorrectly-identified information*), which is comprised of errors and reporting biases, and 3) no information (*hidden information*). One implication of viewing person perception through the lens of the HIDE model is that there are unique insights that one evaluation source (e.g., a judge) might have into a target's characteristics that the other party (e.g., the self) lacks.

The correctly-identified-self component of the model (i.e., self-knowledge) is the knowledge that researchers often aim to capture with self-reports. However, while self-reports predict observable behaviors and important life outcomes remarkably well (e.g., Ozer and Benet-Martinez 2006, Roberts et al. 2007), they are nonetheless vulnerable to errors and reporting biases, so they often capture invalid information. The invalid information piece is described by the incorrectly-identified-self component of the model, which is comprised of both self-deception (i.e., errors) and impression management (i.e., reporting biases). Self-deception refers

to errors in how targets understand themselves. Impression management refers to targets having accurate understanding of themselves but misrepresenting that information to others, usually (but not always) in a positive manner. The incorrectly-identified-self component, therefore, captures both controllable (impression management) and uncontrollable (self-deception) aspects of invalid information. The combination of the correctly-identified-self and incorrectly-identified-self components of the model capture the total information available in a self-report.

What is not included in self-reports is the information captured by the hidden- and hiding-self components of the model. The hidden-self component (self-ignorance) is information that the target is unaware of and therefore does not report. The hiding-self component (self-screening) describes information that the target is aware of but decides not to report. Together, the incorrectly-identified-self, hidden-self, and hiding-self components of the HIDE model capture the information that self-reports cannot accurately assess. Other-reports from judges can, in many circumstances, capture information that is hidden from or incorrectly identified by the self, thus providing insights that self-reports miss.

Paralleling the self-report section of the model, the judge-report section in Figure 1 also shows that information about the target is distributed into three components: 1) valid information that the judge has about the target's characteristics of interest (i.e., correctly-identified-target component; judge-knowledge), 2) invalid information that the judge has about the target because of errors or biases (i.e., incorrectly-identified-target component), and 3) information that judges do know and therefore cannot report (i.e., hidden-target component; judge-ignorance) or that they know but choose not to report (i.e., hiding-target component; judge-screening). The judge-error part of the incorrectly-identified-target component captures information about the target that judges are not able to correctly recognize, whereas the judge-bias part captures the

information that judges are able to correctly recognize but are motivated to misreport in an effort to make the target look better or worse than they actually believe them to be. This might happen after a job interview, for example, when a judge is motivated to make his or her favored candidate look particularly good.

The combination of the correctly-identified-target and incorrectly-identified-target components capture the totality of the information available in the judge-report. The judge-knowledge, judge-error, and judge-bias pieces together reflect how judges view targets and how they represent targets to others. The combination of the incorrectly-identified-target, hidden-target and hiding-target components together capture the information that judges-reports cannot detect accurately.

Relationship of the HIDE model and Existing Interpersonal Perception Models

As a hypothetical example, consider the following. Susan believes that she is highly empathetic. It is true that she understands how others feel and is compassionate in many situations. However, contrary to her belief that she is always highly empathetic toward others, sometimes she tends to ignore others' feelings and can act inconsiderately, especially when she is tired. Susan is completely ignorant about this aspect of herself. Mike, one of Susan's friends, knows that Susan can be inconsiderate. However, in contrast to Susan's self-perception, Mike thinks that Susan is rather inconsiderate in most situations. Mike has no knowledge of the fact that Susan can be highly empathetic in other situations. In this case, how Susan views herself and how Mike views Susan are dissimilar, but each perspective correctly identifies some information about Susan's level of empathy that the other party cannot identify.

The difference between Susan's view of herself and Mike's view of her is closely related to Hogan and Shelter's inner and outer personality (Hogan, 1996; Hogan & Shelter, 1998). Inner

personality is measured by self-reports and captures one's internal motivation and identity. Outer personality, in contrast, is measured by other-reports and captures how the target is viewed by his or her acquaintances based on the target's observable behaviors in social interactions. The HIDE model extends the understanding of self- and other-perceptions by providing the mechanism for understanding when and why one's inner and outer personality converge or diverge from one another. The more that the judge-reports capture knowledge in hidden-self and incorrectly-identified-self components, the more likely the self- and judge-reports are to diverge. In this case, self- and judge-reports provide complementary and non-overlapping information. In contrast, when judge-reports are closely aligned with the correctly identified-self component (self-knowledge) and self-reports are closely aligned with the correctly identified-target component (judge knowledge), self- and judges-reports are more likely to converge.

The HIDE model is distinct from the Johari and SOKA models and the RAM because it does not assume an agreement between different parties to be a prerequisite for validity of both parties in the perception of personality. In these models, the disagreement between self- and other-reports is considered to be an indication of inaccuracy of one of the reporting sources. Besides these models, in the personality literature in general, self-reports have been frequently used as a criterion to validate the accuracy of other-reports, assuming that targets know themselves best. However, the HIDE model does not necessarily interpret the presence of strong agreement as accuracy of both rating sources and neither does it interpret the lack of agreement as inaccuracy of one of the rating sources. For some aspects of personality, targets' self-reports will not be accurate because the information necessary to report on that trait falls into the incorrectly-identified-self and hidden-self sections of the model rather than the correctly-identified-self section. Likewise, while judges' evaluations of targets can be insightful when they

reveal information that is hidden or incorrectly identified by the self, they also can suffer from hidden and incorrectly identified information. Accordingly, the accuracy of self- and other-reports depends on the distribution of personality information detectable by targets and judges. If one party's knowledge about a target is aligned with the other party's incorrectly-identified or hidden information, then the former's evaluation is informative above and beyond the latter's (i.e., incremental validity).

For example, in an extreme hypothetical situation, it is possible that the self-report exclusively measures the correctly-identified-self and the other-report exclusively measures the hidden-self. In this case, self- and other-reports are entirely unrelated (zero correlation), but they both provide valid information about the target, and accordingly both components should relate to observable behaviors. Recalling the hypothetical example of Mike and Susan, it is possible that Susan's self-reported empathy is predictive of how much Susan helps other people who are in need over time, but Mikes' other-reported empathy might predict Susan's inconsiderate behavior toward others when she is tired and not self-aware.

Applying the HIDE model to Character Judgment

Social desirability is critical when considering the relative validity of self-reports versus judge-reports in the HIDE model. Prior research has shown that people often hold biased perceptions of themselves on desirable dimensions (e.g., attractiveness, intelligence; Vazire 2010, Vazire and Mehl 2008). Using the language of the HIDE model, the more desirable the traits of interest, the more self-reports will reflect the incorrectly-identified-self components (i.e., self-deception and impression management). Moral character is an extremely desirable trait—if not the most socially desirable trait— so people have a strong desire to see themselves as moral, leading to self-deception. Moreover, people want to be seen by others as moral, leading to

impression management. Together, self-deception and impression management increase the likelihood that information captured by self-reports will reflect the incorrectly-identified-self component. Consequently, judge-reports could complement or replace self-reports to the extent that they tap into valid information that reflects the incorrectly-identified-self. Moreover, judge-reports could capture information in the hidden-self (i.e., self-ignorance) and hiding-self (i.e., self-screening) components that are not accessible to the individuals providing self-reports.

Judge-reports of moral character can be provided by people who know the target well (i.e., well-acquainted others) or by strangers who have no relationship with the target but nonetheless have access to information about their moral character. We often assume that well-acquainted others will be better judges than strangers, and studies generally support this claim (Funder 1995, Kenny et al. 1994). However, the HIDE model suggests that, in some circumstances, evaluations made by strangers can be more informative than those provided by well-acquainted others, even though the latter have the opportunity to observe targets in various situations over time. Strangers are likely to be more accurate than friends when friendship hinders the ability to correctly construe targets' moral character. Using the language of the HIDE model, judgments by strangers are likely to be more accurate than judgments from well-acquainted others (e.g., friends) in circumstances in which the latter have incorrectly-identified-target knowledge (i.e., judge-error or judge-bias). Consequently, it is necessary to develop a tool that judges can use to accurately extract information concerning the moral character of strangers.

Balance theory (Heider, 1957; Insko, 1981) explains why people tend to perceive or believe good things about their friends and bad things about their enemies, and thus provides one explanation why well-acquainted others' moral character evaluations about targets may be located in the judge-error and judge-bias zones of the HIDE model. According to the theory,

individuals' perceptions of others depend on the social relationships these individuals share (Insko, 1981). When a judge has a positive relationship with a target (e.g., friendship), the judge tends to ascribe high value to the target on positive traits, but lower value on negative traits. This pattern achieves balance (i.e., consistency) between the positive "unit relationship" of having a friendship with the target and the positive evaluations people have of morality. In other words, the following three cognitions are balanced: This person is my friend (+); I value morality (+); my friend is moral (+). Imbalance in this triad of cognitions leads to cognitive dissonance and motivation to reduce the inconsistency. The consistency motive described by balance theory thus explains why judges might misconstrue targets to be consistent with their existing relationship with them, otherwise the judges would feel discomfort from inconsistency. Well-acquainted others' evaluations are therefore susceptible to conscious or unconscious bias in evaluations. However, by definition, strangers do not have relationships with targets, and thus their evaluations should be less likely to be pushed into the judge-bias and judge-error zones of the HIDE model.

Therefore, reducing the hidden-target zone of the model will be an important condition that would allow strangers to form accurate moral evaluations of targets. Moreover, for the judge-reports to complement or replace self-reports, the correctly-identified-target component should include knowledge contained in the hidden-self and/or incorrectly-identified-self components of the HIDE model. It follows, then, that it is necessary to develop a tool that judges can use to accurately extract information about moral character traits that the targets themselves are unaware of and/or less able to control.

Moral Character Judgment via Written Responses to Job Interview Questions

An interesting and practical tool that judges might use to evaluate strangers' moral character is to ask open-ended questions designed to reveal the "hidden" aspects of unethical tendencies–those that job applicants are unaware of and less able to control. To test the plausibility of this claim I have developed a battery of interview questions that covertly elicit peoples' unethical tendencies through their spontaneous written responses.

I focus on written responses for several reasons. Most importantly, previous studies have shown that performing an expressive task (i.e., writing) requires an individual to engage in impromptu thinking, and the dispositions reflected in such expressions are difficult to counterfeit (Hojbotă 2015). Second, evaluations based on written responses (compared to other media, such as face-to-face conversations) can help reduce certain factor that might bias judges (e.g., the attractiveness of candidates; Cann et al. 1981).

The interview questions developed in this study are presented in Appendix 1. The questions were modeled after behavioral interview questions commonly employed in research and practice (Blackman 2002, Hoevemeyer 2005). Each interview question is developed to reveal aspects of traits diagnostic of unethical tendencies. What targets talk about (e.g., past events that are salient to them), whether they consider others' needs in difficult situations, and how they feel when their behaviors might influence others (e.g., feeling guilty when their behaviors negatively influence others) is likely to provide judges with explicit and implicit information that could enable them to make accurate moral character judgments. For example, the "Mistake question" asks job applicants to recall a mistake they made at work and to report how they felt and behaved at the time. Prior research has shown that unethical individuals experience less guilt following wrongdoing (e.g., Cohen et al. 2016). Although these individuals may not overtly admit it, their responses to this question reveal that they elaborate much less on

past experiences of guilt following a mistake, and this response pattern makes it possible to identify them.

The Wisdom of Crowds in the Evaluation of Interview Responses

In evaluating people using the interview method, we need to consider the possibility that inter-rater reliability might be low. Indeed, Previous research has shown that inter-rater reliability for evaluating interviews is generally low because different interviewers often apply different standards when evaluating applicants (Arvey and Campion 1982, Highhouse 2008). However, I propose that while each individual might not be able to judge targets reliably and accurately, a group of judges could do so. By recruiting a large number of heterogeneous judges, the biases and errors stemming from their individual idiosyncrasies can be offset.

Consistent with this reasoning, research on the "wisdom of crowds" shows that collectives composed of independent judges often make more accurate judgments and decisions than do solo individuals (Davis-Stober et al. 2014, Larrick and Soll 2006, Mannes 2009). The "wisdom of crowds" is based on the premise that the aggregate of multiple independent judgments will be more reliable because high and low errors offset each other. For example, a very positive or lenient judge who rates all candidates highly will be offset by a very negative or conservative judge who rates all candidates poorly. However, having a large "crowd" of judges often entails substantial costs (of time, money, etc.). Hence, knowing how many judges are required to obtain reliable judgments and predictions of unethical behaviors is critical for optimizing selection procedures.

CHAPTER II

The Collective Wisdom in Forecasting Unethical Behavior

In Chapter 2, I investigate whether groups of naïve judges can predict others' unethical behaviors by evaluating their moral character from written responses to the interview questions designed to elicit information about people's implicit aspects of moral character. My prediction is that impromptu thinking and language usage captured in written responses to these questions reveal information about targets' moral character that judges can use to make valid character judgments. Performing an expressive task (i.e., writing) requires an individual to engage in impromptu thinking, and dispositions reflected in such expressions are difficult to counterfeit (Hojbotă, 2015). For example, what targets talk about (e.g., past events that are salient to them), whether they consider others' needs in difficult situations, and how they feel when their behaviors might influence others (e.g., feeling guilty when their behaviors negatively influence others) are likely to provide judges with information that could enable them to make valid moral character judgments. In Chapter 2, I examine the validity of moral character judgments based on targets' written responses by measuring how well they predict unethical behaviors.

With regard to the HIDE model discussed in Chapter 1, what the studies in this chapter test are whether the information captured in targets' written responses to behavioral interview questions provide judges with correctly-identified-target knowledge. However, given the potentially low observability of moral character information, it is possible that information relevant to judging moral character remains hidden, resulting in judge-ignorance rather than judge-knowledge. Because the judges in these studies do not know the targets, I assume that incorrectly-identified-target knowledge (i.e., judge-error and judge-bias) is relatively inconsequential. Thus, the focal comparison in this chapter is between correctly-identified-target knowledge and hidden-target knowledge. Predictive validity of unethical conduct provides initial

evidence that information captured in targets' written responses to behavioral interview questions provide judges with correctly-identified-target knowledge.

Chapter 2 consists of three empirical studies examining the wisdom of crowds in forecasting unethical behaviors using this text-based interview method. In studies 1 and 2, I crowd-sourced large sets of judges online and these judges evaluated targets' moral character from written interview responses. Study 3 extended the findings of studies 1 and 2 by determining the judge size at which the crowd effect occurred when forecasting unethical behavior using the text-based interview method.

Study 1

In Study 1, I investigated the predictive validity of judges' evaluations in a laboratory experiment in which target participants had the opportunity to over-report their performance on a problem-solving task to earn additional money. I examined whether the aggregated evaluations of multiple judges, formed from written responses to the interview questions, predict how frequently targets engage in cheating.

Method

First, two behavior-based interview questions were developed to extract targets' moral character information. The questions were modeled after behavioral interview questions commonly employed in research and practice (Blackman, 2002; Hoevemeyer, 2005):

- Please tell us about a time when you made a mistake at work. How did you feel when this occurred? What did you do? What, if anything, did you learn from this experience? [*Mistake*]

- Please describe an experience in which you were faced with a difficult dilemma at your job—a situation where you found it hard to decide what to do. What factors did you consider? What did you do? What, if anything, did you learn from this experience? [*Dilemma*]

Each interview question is developed to reveal aspects of traits diagnostic of unethical tendencies. The mistake question asks job applicants to recall a mistake they made at work and to report how they felt and behaved at the time. Prior research has shown that unethical individuals experience less guilt following wrongdoing (e.g., Cohen et al. 2016). Although these individuals may not overtly admit it, their responses to this question reveal that they elaborate much less on past experiences of guilt following a mistake, and this response pattern makes it possible to identify them. The dilemma question gives targets the opportunity to reveal the extent to which they are considerate of others and mindful of how their decisions and actions affect other people. We designed this question because we assumed that high-moral-character targets would be more likely than low-moral-character targets to mention such considerations. Each target responded to one of these two questions, after reading the following instructions.

> *Imagine that you have been selected to interview for your dream job. The employers want to conduct an online interview before you meet them face to face. You will be asked questions about yourself and past experiences you may have had. Please use real examples from your life when responding. Please do not include last names or any other personally identifiable information in your response. Remember: you need to answer the following questions honestly, but in a way that makes you look like the best possible job candidate.*

Data Collection from Targets

The targets who responded to the interview questions in this study were 195 U.S. adults who participated in an experiment in a mobile research laboratory parked in the city of Pittsburgh, Pennsylvania. In addition to answering one of the two interview questions, participants completed a problem-solving task in which they had the opportunity to lie about their performance, and a computerized survey in which they answered the five-item guilt

proneness scale (GP-5; Cohen, Kim, & Panter, 2014), the HEXACO-60 personality inventory (Ashton & Lee, 2009), and questions capturing demographic information.[1]

The problem-solving task was based on methods used by Shu, Mazar, Gino, Ariely, and Bazerman (2012). Participants were given a worksheet containing 20 matrices with 12 three-digit numbers within each matrix. They had five minutes to find two numbers in each matrix that added to 10.00. Each correctly identified pair of numbers was worth $0.25 in earnings, for a maximum bonus payment of $5.00. Participants learned that they would work on the task for five minutes and then would be asked to calculate the number of problems they solved correctly and indicate this number and how much money they should be paid on a payment form, after they had recycled the matrices worksheet. Unbeknownst to the participants, we were able to link each participant's problem-solving performance to his or her payment form by a three-digit identifier contained in each of the documents. One three-digit number in the bottom matrix on the problem-solving worksheet was identical to three digits in the payment form number. At the end of each day of data collection we collected all the matrices worksheets from the recycle bin and compared each participant's reported performance on the payment form to his or her actual performance on the worksheet. Participants were considered to have cheated when the number of problems they reported solving was greater than the number they actually solved correctly on the worksheet.

After participants worked on the problem-solving task for five minutes, they put their worksheets in the recycle bin, and wrote down the number they solved correctly and how much money they earned on the payment form. Then participants completed the computerized survey that included a question asking them to describe themselves, one of two questions (either the

[1] Two additional participants completed the study but were excluded from the analyses because they answered 19 out of 20 items correctly on the problem-solving task, and therefore had little opportunity to cheat compared to other participants.

Dilemma or the Mistake question), the GP-5, the HEXACO, and demographic questions. Following the computerized survey, participants handed their payment forms to the experimenter, were paid according to the number of problems they indicated solving on the payment form, and were provided with a debriefing form that explained that the true purpose of the study was to examine cheating.

Data Collection from Judges

One hundred and two participants were recruited from a university-administered subject pool to complete a web-based study, in which they judged the targets' moral character (55.9% were female; the average age was 21.6, ranging from 18 to 69) from Study 1. They were given class credit for their participation. Each judge rated interview responses from 20 randomly selected targets. Each interview response was rated by an average of 15 judges. Judges read the following instructions:

> *In making your judgment of moral character, please consider the following definition.*
>
> *Moral character is a term used to describe an individual's disposition to think, feel, and behave in an ethical manner. People with high levels of moral character consider the needs and interests of others, and how their own behavior affects other people. When they do something wrong they feel guilty and try to correct for what they did, even if no one knows about it. In general, those with high moral character are benevolent, trustworthy, and compassionate. In contrast, people with low levels of moral character are callous, manipulative, and more focused on themselves than on other people. When they do something wrong they are unlikely to feel bad about their behavior or attempt to correct for their mistakes. In general, those with low moral character are cruel, dishonest, and inconsiderate.*

Each judge rated moral character by responding to the question: Do you consider the author of this response to be a moral person? *[1 (Extremely weak moral character), 2 (Weak moral character), 3 (Neither weak nor strong), 4 (Strong moral character), 5 (Extremely strong moral character)]*.

Results

The criterion variable, cheating, is operationalized as the number of matrices the participants claimed they solved minus the number they actually solved correctly. The descriptive statistics and correlations among targets' cheating frequencies, self-reported moral character traits, and judges' average-moral-character-rating are presented in Tables 1 and 2. I found the negative relationship emerged between targets' self-reported Conscientiousness and their frequency of cheating. Honesty-Humility and guilt proneness did not show statistically significant correlations; nevertheless, the directions of their relationships were consistently negative. Figures 2 and 3 depict the relationship between judges' average-moral-character-ratings and the extent to which targets cheated on the problem-solving task in the Mistake and Dilemma question conditions.

I formally tested the predictive validity of judges' average-moral-character-ratings by conducting negative binomial regression analyses. In each analysis, the number of correctly solved matrices was controlled because participants who solved more matrices correctly had less opportunity to cheat. In total, three different sets of analyses were conducted. The results were similar, regardless of whether the Mistake and Dilemma questions were analyzed together or separately. The results from the separate analysis for each question are presented in Table 3.

In the first model, only the judge-reports were entered. The results indicated that judge-reported moral character negatively and significantly predicted the extent to which targets cheated in the problem-solving task, regardless of whether those ratings were made from targets' written interview responses to the Mistake or Dilemma question. In the second model, only the self-reports were entered. Self-reported Conscientiousness negatively, and marginally significantly, predicted the extent to which targets cheated in the Mistake question condition.

Finally, in the third model, targets' frequency of cheating was regressed on both judge- and self-reports to test which rating source is more predictive.

The results indicated that only the judge-reports had incremental validity, which means that the judge-reports were more informative than the self-reports in predicting cheating. The net effects of judges' moral character judgments were negative and significant for the Mistake question condition and negative and marginally significant for the Dilemma question condition.

Discussion

Although the criterion I used to measure unethical behavior in Study 1—lying about one's performance on a laboratory task—has strong internal validity and was directly observable to the experimenter (as opposed to self-reported), it lacks external validity. The specific form of cheating we examined and the laboratory context in which it occurred do not correspond to the kinds of cheating that occur in real-life settings. Therefore, investigating the predictive validity of this text-based interview method in real social interaction settings would increase the generalizability of the laboratory findings.

Study 2

Although the criterion I used to measure unethical behavior in Study 1—lying about one's performance on a laboratory task—has strong internal validity, it lacks external validity. In Study 2, I investigated the predictive validity of this written-interview-response method using a field study of working adults with counterproductive work behavior (CWB) as a criterion. CWB, also known as workplace deviance, CWB is defined as employees' volitional behaviors that harm or intend to harm the people in an organization and the organization itself and is perceived as unethical by employees in general (Cohen et al., 2014). CWB includes a wide range of

unethical work behaviors, such as falsification of expense reports, stealing, and interpersonal abuse.

Method

Interview Questions

In addition to the Mistake and Dilemma questions, Study 3 used an additional, following question:

- How would your current or last employer describe you? *[Employer]*

I reasoned that targets' assessments of their employer's perceptions about them might be indicative of targets' humility, with high-moral-character targets being more modest and unassuming compared to low-moral-character targets.

Data Collection from Targets

The target participants in Study 3 were 495 employed U.S adults recruited by an online survey firm (Qualtrics)[2]. These target participants were randomly assigned to answer one of the interview questions. Employees' CWB was measured using the 32-item inventory developed by Spector and his colleagues (2006). Finally, participants were administered the HEXACO-60 revised personality inventory (Ashton & Lee, 2009) and the five-item guilt proneness scale (GP-5; Cohen et al., 2015).

Data Collection from Judges

In total, 677 U.S. residents were recruited via Amazon's Mechanical Turk website (www.mturk.com). Eligible participants were those with an at least 90% approval rating on previous tasks. We excluded five participants who did not complete the study or who failed one or more attention checks embedded in the survey, leaving a final sample of 672 participants.

[2] These participants are a subset of participants in a larger project investigating with larger number of interview questions. The results for other questions are available from the author.

Overall, 52% were female and their average age was 36.98 years (Range: 18-83). Each participant rated interview responses from 20 randomly selected targets. Each interview response was rated by an average of 17 judges. The rating instructions and definitions of moral character traits were the same ones used in Study 1.

Results

The descriptive statistics for self-reported CWB, Honesty-Humility, Conscientiousness, and guilt proneness are presented in Table 4. Consistent with previous research, self-reported Honesty-Humility, Conscientiousness, and guilt proneness showed negative relationships with CWB (Cohen et al., 2013). The descriptive statistics for judge-reported moral character across the thee interview questions are presented in Table 5. Across all interview questions, judges' moral character evaluations negative predicted CWB.

The predictive validity of moral character judgments was tested using negative-binomial analyses. For each interview question condition, three sets of analyses were conducted. The first set of analyses examined the predictive validity of judge-reported moral character while the second set of analyses examined the predictive validity of self-reported traits. Finally, in the third set of analyses, both self- and judge-reports were entered simultaneously. The results are presented in Table 6. The results indicate that, across all five conditions, judge-reported moral character negatively and significantly predicted the frequency with which targets engaged in CWB. Self-reported Honesty-Humility, Conscientiousness, and guilt proneness also negatively and significantly predicted CWB. Finally, when judge- and self-reported moral character traits were entered simultaneously, only self-reports provided incremental validity. However, CWB was measured with self-reports, which is influenced by the method bias (i.e., shared variance). Nonetheless, judges' moral character ratings, while not significant at the standard $\alpha < .05$ level,

showed the expected negative patterns for all interview question conditions, and they were marginally significant for the Employer question condition.

Discussion

In Study 2, I found that judges' average-moral-character-rating have predictive validity with workplace deviance as the criterion. It is a well-established fact that CWB is ubiquitous in organizations, causes organizations substantial economic damages, and hurts individuals and society (Bennett & Robinson, 2000; Budd, Arvey, & Lawless, 1996; Glomb, 2002). The findings of Study 2 suggest that CWB can be reduced greatly by identifying job candidates and employees who are low in moral character using the interview questions developed in this research and monitoring these employees closely to prevent CWB from occurring.

Study 3

In Studies 1 and 2, a large number of participants were recruited to play the role of judges. Using multiple judges increases the reliability of aggregated evaluations. However, it also entails substantial costs (of time, money, etc.). Indeed, most organizations employ relatively small groups of interviewers to evaluate job candidates. Therefore, it is important that we be able to determine the minimum number of judges required to form reliable character judgments using this written-interview-response method. To do so, I used Generalizability (G) theory (Cronbach et al. 1963) to calculate the changes in inter-rater reliability as the number of judges varies. However, G theory analyses require that the same set of targets be evaluated by the same set of judges. In Studies 1 and 2, calculating inter-rater reliability was not possible because judges were randomly assigned to different sets of targets. In Study 3, however, six judges read and evaluated the entire set of targets, thus allowing me to conduct Generalizability (G) theory analyses.

Method

Six undergraduate research assistants were recruited and read the entire set of interview responses from Study 1, then rated each target's overall moral character, specific moral character traits, and other characteristics. Each judge indicated their rating of overall moral character by responding to the question: *Do you consider the author of this response to be a moral person? [1 (Extremely weak moral character), 2 (Weak moral character), 3 (Neither weak nor strong), 4 (Strong moral character), 5 (Extremely strong moral character)*]. No specific definition or criteria for evaluating moral character was provided to the judges. However, each judge also made a number of other ratings of the targets, which may have influenced their judgment of moral character. Specifically, prior to judging each target's overall moral character, the judges were given definitions of *Guilt Proneness, Conscientiousness, Honesty-Humility,* and *Agreeableness*[3], and were asked to rate each target on these traits relative to a typical job applicant (ranging from extremely low to extremely high) [4].

The present study focuses on judges' responses to the global moral character question (*Do you consider the author of this response to be a moral person?*), and three specific traits (Honesty-Humility, Conscientiousness, and guilt proneness) that have been identified as key moral character traits (Kim & Cohen, 2015).

Results & Discussion

The descriptive statistics of the six judges' average ratings of moral character, Honesty-Humility, Conscientiousness, and guilt proneness, are presented in Table 7. Both Study 1 judges

[3] Definitions used in Study 3 were also used in Study 4. These definitions are presented in the Method section of Study 4.

[4] In addition, after indicating their judgment of moral character, each judge answered three additional questions related to moral character: *Do you think this person considers the needs and interests of others, and how his/her own actions affect other people?*; *Do you think this person values morality and wants to see himself or herself as a moral person?*; and *This person participated in a laboratory experiment in which they could cheat by over-reporting their performance in a problem-solving task to earn money. Do you think this person cheated in the experiment? [No, this person was honest (did not cheat at all); Yes, this person cheated a little; or Yes, this person cheated a lot]*. These variables were measured for other purposes, so it is not reported in this manuscript. However, the results are available from the author.

and Study 3 judges evaluated the moral character of Study 1 targets and the correlation between these two sets of judges was strong ($r = .80$, $p < .001$). The correlations between cheating frequency, judge-reported traits, and self-reported traits are presented in Table 8. The results indicated that judges' ratings of moral character were more likely to predict targets' cheating frequency than self-reported Honesty-Humility, Conscientiousness, and guilt proneness.

To determine the number of judges required to reliably evaluate targets' moral character traits, I used Generalizability Theory (Cronbach, Nageswari, & Gleser, 1963), which enables us to estimate how the reliability of judgments varies with the number of judges. The results of these analyses are presented in Table 9. Both interview questions showed high levels of consensus, such that the current six judges had greater than .70 reliability. Increasing the number of judges becomes decreasingly beneficial as the number of judges increases.

To formally test the predictive validity of the moral character judgments with small groups of naïve judges, negative binomial regression analyses were conducted for each interview question. For each question condition, two different sets of analyses were conducted (see Table 6). In the first set of analyses, target cheating frequency was regressed on the average-moral-character-rating of six judges. In the second set of analyses, target cheating frequency was regressed on moral character judgments from each judge *individually* to investigate the possibility that each individual was able to detect target's moral character. Replicating Studies 1 and 2, the first set of analyses found that six judges' average-moral-character-ratings significantly and negatively predicted targets' unethical behavior, and this was true for both interview question conditions. The second set of analyses provided partial support for individual-level accuracy in judging strangers' moral character based on written interview responses. Every judge's moral character judgments significantly and negatively predicted target cheating

frequency in the Mistake question condition. However, the predictive validity of individual-level moral character judgments was weaker and less robust for the Dilemma question condition. Although somewhat inconsistent, the individual level accuracy observed in this study is inspiring, considering the limited information provided to judges (a brief paragraph consisting of an average of 76.21 words).

Discussion

The most striking and interesting findings of Study 3 is that even a very small number of judges (i.e., six judges) could reliably estimate targets' moral character. I also compared individual-level prediction with the collective, aggregate-level prediction and showed that the predictive validity (i.e., effect size) was much higher for the latter.

This finding demonstrates that the "wisdom of crowds" phenomenon (i.e., that the quality of human judgment increases as the number of judges increases) also applies to moral character judgments, such that collectives of individuals detected strangers' moral character more accurately than individuals did alone (Larrick, Mannes, & Soll, 2012). This phenomenon has important practical implications for organizational contexts. In interview settings, for example, an interviewer might not be able to detect moral character accurately by him/herself, but a small set of independent interviewers (e.g., six judges) might be able to.

General Discussion

According to the HIDE model of moral character, judges who do not know the targets might be able to capture aspects of targets' moral character that self-reports do not capture. In this chapter, I examined this theoretical prediction by evaluating how well judges' aggregated moral character ratings predict targets' unethical behaviors. In Study 1, I conducted a laboratory experiment in which target participants had the opportunity to over-report their performance on a problem-solving task to earn additional money. I found that judges' average-moral-character-

rating significantly predicted the extent to which targets cheated on the problem-solving task. In Study 2, I replicated this finding with a different criterion, CWB, which includes a wide range of harmful work behaviors, such as falsification of expense reports, stealing, absenteeism, and interpersonal abuse. In line with the study 1, I found that judges' average-moral-character-ratings significantly predicted the frequency of which targets reported engaging in CWB. In Study 3, I found that even a very small number of judges (i.e., six judges) could reliably estimate targets' moral character.

Moral character judgment is probably the most important interpersonal judgment. If we can detect strangers' moral character, it would have important practical applications in selection and promotion contexts within organizations, as well as important theoretical implications for understanding how we come to know individuals, and specifically whether they are likely to behave ethically. The most significant contribution of Chapter 2 is that it supports the notion that moral character can be detected in zero-acquaintance settings in which the targets provide only limited personal information about themselves to the judges.

CHAPTER III

Moral Character Information Captured by Written Interview Responses

Chapter 2 demonstrated that judges' evaluations of target individuals' moral character from their written responses predicted the targets' unethical behaviors. In Chapter 3, I investigate how to increase the predictive power of this text-based interview method with regard to unethical behavior. Each interview question is designed to capture different aspects of moral character; thus, one way to improve predictive power would be to evaluate the targets on more narrowly defined dimensions that are matched to information revealed by their answers to each interview question.

For example, the mistake question might reveal targets' guilt proneness or Conscientiousness. This Mistake question asked job applicants to recall a mistake they made at work and to report how they felt and behaved at the time. Prior research has shown that unethical individuals are less likely to feel guilty after wrongdoing. Although unethical individuals may not admit overtly to this lack of guilt, their responses to the Mistake question might reveal that they elaborate much less compared to other respondents on past experiences of guilt following a mistake, which makes it possible for judges to evaluate targets' guilt proneness with accuracy. Additionally, answers to the Mistake question might reveal targets' Conscientiousness because highly conscientious individuals are more likely to expend effort to correct for their mistakes and thus may elaborate on what they did and what they learned from their past mistakes.

The Dilemma and the Employer questions were designed to capture targets' Honesty-Humility. For example, persons who are high in modesty (i.e., not narcissistic) and generous to others (i.e., high in greed-avoidance) might talk about how their decisions influenced others rather than focusing on themselves in answering the dilemma question. In response to the

Employer question, targets who are more modest and humble may be less likely to assume that their employers discussed only extremely positive elements in describing targets.

To determine which aspects of targets' moral characters were revealed in written responses to each interview question, the judges evaluated respondents on three narrowly and distinctively defined moral character traits: Honesty-Humility, Conscientiousness, and guilt proneness. I focused on these three dimensions for several important reasons. First, the HIDE model compares self-reports and judge-created reports, and the validity of these three traits is strongly supported by the rich literature of personality traits. Moreover, there are already well-established tools for self-reporting and peer-reporting for these traits. It is critical to note that self-reported ratings for these traits as well as ratings provided by others who are well-acquainted with the targets have been shown to predict targets' unethical behaviors across diverse situations (Kim & Cohen, 2015).

Chapter 4 consists of three empirical studies. In Study 4, the research question, which explores what aspects of targets' moral character are conveyed in written interview responses—is investigated via the convergent and divergent validity of judges' evaluations of Honesty-Humility, Conscientiousness, and guilt proneness. Studies 4 and 5 answer the research question discussed in Chapter 3 by investigating the predictive validity of unethical behavior as measured in a laboratory setting and reported by targets' peers in the work setting.

Study 4

Study 4 examines to what extent targets' written responses to the Mistake, Dilemma, and the Employer interview questions reveal Honesty-Humility, Conscientiousness, and guilt proneness by investigating the convergent and divergent validity of judges' ratings for these three dimensions.

Data Collection from Targets

The targets who provided the responses to the interview questions were 406 U.S. adults recruited from Amazon Mechanical Turk. Each of these participants (i.e., targets) answered three randomly chosen interview questions out of five[5]. In this study, I focus on responses to three questions, which are the Mistake, the Dilemma, and the Employer questions. Targets who responded to the interview questions with fewer than 20 words were excluded in the current study because such short responses would not provide enough information for raters to make personality judgments. Following the open-ended interview questions, the targets answered several personality questionnaires, including Ten Item Personality Measure (TIP; Gosling, Rentfrow, & Swann, 2003) and GP 5 (Cohen et al., 2014). Similar to how Conscientiousness and other personality traits were measured in TIPI, Honesty-Humility was measured with two pairs of traits: "honesty, fair", "boastful, greedy."

Data Collection from Judges

Five undergraduate research assistants read the entire set of interview responses and rated each target's Honesty-Humility, Conscientiousness, Agreeableness, and guilt-proneness. Agreeableness is not a key indicator of moral character trait in the HEXACO model (Kim & Cohen, 2017) and is included in the current study as a comparison evaluation dimension. The judges were asked: Compared to a typical job applicant, do you consider the author of this essay to be low or high on [Honesty-Humility, Conscientiousness, Agreeableness, and guilt-proneness]? They could endorse: 1 (Extremely Low), 2 (Low), 3 (Neither Low nor High), 4 (High), 5 (Extremely High). Judges read the following instructions for each trait.

> ***Guilt Proneness****: Guilt proneness is a personality trait indicative of a disposition toward experiencing negative feelings about personal wrongdoing, even when the wrongdoing is private. In judging guilt proneness, think about whether the person would feel bad about*

[5] The results of the other two questions are presented in the Appendix.

making a mistake or committing a transgression even if no one knew about what they did. A person high on Guilt Proneness feels bad about their behavior when they do something wrong; a person low on Guilt Proneness does not feel guilty about wrongdoing.

***Conscientiousness**: Conscientiousness is a personality trait indicative of a disposition toward organization, diligence, perfectionism, and prudence. In judging Conscientiousness, think about whether the person is hard-working, careful, and thorough when working or completing tasks. A person high on Conscientiousness is dependable and self-disciplined; a person low on Conscientiousness is disorganized and careless.*

***Honesty-Humility**: Honesty-Humility is a personality trait indicative of a disposition toward fairness, sincerity, modesty, and greed-avoidance. In judging Honesty-Humility, think about whether the person is truthful and humble in their interactions with others. A person high on Honesty-Humility is honest and fair; a person low on Honesty-Humility is boastful and greedy.*

***Agreeableness**: Agreeableness is a personality trait indicative of a person's forgivingness, gentleness, flexibility, and patience. In judging Agreeableness, think about whether the person is tolerant and peaceful in their interactions with others. A person high on Agreeableness is sympathetic and warm; a person low on Agreeableness is critical and quarrelsome.*

Results

The descriptive statistics and correlations among targets' self-reported and judge-reported Honesty-Humility, Conscientiousness, guilt proneness, and Agreeableness are presented in Tables 11 and 13.

I calculated two types of convergent validity indices to examine the extent to which information about targets' Honesty-Humility, Conscientiousness, guilt proneness, and Agreeableness were revealed in their responses to each interview question. First, I used the Generalizability Theory (Cronbach, Nageswari, & Gleser, 1963) to calculate inter-judge reliability by varying the number of judges. If different judges perceive a particular target's traits in dissimilar ways, it is difficult to argue that the judges' ratings provide unique and consistent information. This lack of consensus on targets' traits can reflect a lack of information about targets with respect to that trait. The results of Generalizability theory analyses are presented in Table 14. Analyses revealed that judges' evaluations of Honesty-Humility had the lowest levels

of consensus across all interview question conditions. The employer question condition, in contrast, resulted in stronger levels of consensus for the Conscientiousness dimension compared to the Mistake and Dilemma question conditions.

Second, I calculated correlations between target-reports and judge-reports (see Table 15). In the HIDE model, it is theoretically possible for self-reports and judge reports to capture entirely non-overlapping aspects of a trait (i.e., zero correlation) but both are still valid. This can happen when judges' correctly-identified-target component captures only the targets' self-ignorance (see Figure 1). However, this is an extreme scenario. The results presented in Chapter 1 suggest that this did not happen for the moral character judgments. In Studies 1, 2, and 3, the self-reports and judge-reports both were predictive of targets' unethical behavior. More importantly, judges' moral character evaluation and self-reports of Honesty-Humility, Conscientiousness, and guilt proneness were correlated positively, meaning that self-reported and judge-reported moral character assessments tapped into overlapping information. Therefore, the positive and significant correlations between the self- and judge-reports provided convergent validity evidence of judges' ratings on the evaluation dimension in question. Further, the lack of positive correlations among other evaluation dimensions in that same question provided evidence of divergent validity. These results are presented in Table 15.

The Mistake question resulted in the strongest positive correlations for Conscientiousness evaluations and the second best for the Honesty-Humility. The dilemma and employer questions resulted in good self-judge agreements on guilt proneness.

Finally, although judges' ratings for Agreeableness revealed strong consensus across questions, they did not reveal any significant agreements in terms of self-judge convergence. These results are consistent with expectations because those interview questions were designed

to capture targets' tendency to think, feel, and behave ethically, whereas Agreeableness in the HEXACO framework is not related to such characteristics. Therefore, the lack of self-judge correlations regarding Agreeableness provided divergent validity evidence for the interview questions developed in this research.

Discussion

The results of Study 4 suggest that targets' written responses to the Mistake question revealed a significant amount of information about Conscientiousness. Originally, I reasoned that the mistake question could diagnose targets' guilt proneness given that people who are high in guilt proneness might report that they felt bad after making a mistake. The results of Study 4, however, suggested that targets' guilt proneness was not revealed effectively by responses to the mistake question. Instead, targets' guilt proneness was better measured based on judge reports from the dilemma question. Judges' ratings of guilt proneness had the strongest levels of consensus among guilt proneness evaluations compared to other interview questions and had a positive self–judge correlation.

Additionally, it is noteworthy that judges' ratings of Honesty-Humility had the lowest levels of consensus among all interview question conditions. Moreover, although the Employer question was designed to elicit information about targets' Honesty-Humility, inter-judge reliability regarding the Honesty-Humility dimension of the Employer questions was low. Instead, the Employer question resulted in strong consensus in terms of the Conscientiousness evaluation.

Finally, it is important to note that judges' ratings for Honesty-Humility had the lowest levels of consensus among all questions for several possible reasons. It is possible that targets' Honesty-Humility levels were not revealed effectively by any of the three interview questions

used in Study 4. In addition, the scope of the Honesty-Humility factor was too broad for judges to evaluate it consistently. This factor comprised four distinctive elements: fairness, sincerity, greed-avoidance, and modesty. In the HEXACO framework, self-reported ratings for four specific elements of Honesty-Humility (fairness, sincerity, greed-avoidance, and modesty) form one global factor: the Honesty-Humility factor. In other words, when measured by self-reports, fairness, sincerity, greed-avoidance, and modesty share a strong variance, which is interpreted as the Honesty-Humility factor. However, it is possible that judges' ratings of those four elements were not homogeneous, so combining these four elements into one overarching factor may not be worthwhile. Finally, it is possible that judges' ratings of fairness, sincerity, greed-avoidance, and modesty are not all valid. Because these evaluations may be erroneous, the Honesty-Humility judgments were not consistent across judges. Chapter 4 of this dissertation examines these possibilities.

Study 5

Study 4 examined whether targets' Honesty-Humility, Conscientiousness, and guilt proneness were revealed by different interview questions by investigating the convergent and divergent validities of judges' ratings of these dimensions based on targets' written responses to each question. Study 5 examined the relative predictive validity of judges' ratings of Honesty-Humility, Conscientiousness, and guilt proneness compared to their ratings of moral character. If only certain aspects of moral character (e.g., Conscientiousness) could be obtained from each interview question, then judges' average ratings on smaller, matching dimensions (e.g., the Conscientiousness evaluation of responses to the Mistake question) would be more valid than judges' average ratings on a larger scope (i.e., the moral character evaluation). Therefore, I compared how well judges' ratings of n Honesty-Humility, Conscientiousness, and guilt proneness predicted targets' unethical behaviors in different interview questions with how well

judges' ratings of moral character predicted unethical behavior. Based on the findings of Study 4, I hypothesized that judges' average rating for Conscientiousness was more predictive of targets' unethical behavior than judges' average rating of moral character in the Mistake question. Additionally, I hypothesized that judges' average rating of guilt proneness was more predictive of targets' unethical behavior than the judges' rating of moral character in the dilemma question.

In addition to examining the relative predictive validity of moral character versus Honesty-Humility, Conscientiousness, and guilt proneness dimensions, Study 5 investigated the relative predictive validity of self-reports versus judges' reports for these three dimensions. The HIDE model predicts that judges' average ratings for moral character based on targets' written interview responses should be more valid than ratings of moral character provided by the targets themselves. This prediction was supported in Studies 1 and 3; however, as discussed in Chapter 2, when comparing self- and judge-reported moral character traits, the evaluation scopes were different. Judges evaluated on a larger scope (i.e., moral character) compared to the dimensions of self-reports (i.e., Honesty-Humility, Conscientiousness, and guilt proneness). In Study 5, I compared the predictive validity of targets' self-reports and judge reports based on the same dimensions (i.e., Honesty-Humility, Conscientiousness, and guilt proneness). Considering that judge-reports were more valid in terms of evaluation dimensions matching to interview questions, I hypothesized that judges' average ratings for Conscientiousness would be more predictive of unethical behavior than self-reported Conscientiousness. I also hypothesized that judges' average rating for guilt proneness would be more predictive of unethical behavior than self-reported guilt proneness.

Method

In total, 500 participants were recruited from an online participant pool to read and evaluate the interview responses from Study 1. The targets' frequency of cheating, as measured in Study 1, was used as a criterion to determine the relative predictive power of unethical behavior in judges' ratings of different dimensions (moral character vs. Honesty-Humility, Conscientiousness, and guilt proneness) and different rating sources (self vs, judge).

This study used 2 by 3 between-conditions design. Judges were randomly assigned to one of two interview question conditions (i.e., Mistake, Dilemma) and one of three evaluation dimension conditions (i.e., Honesty-Humility, Conscientiousness, guilt proneness). The rating instructions and definitions for Honesty-Humility, Conscientiousness, and guilt proneness were the same ones used in Study 4. Each judge rated interview responses from 20 randomly selected targets. Each interview response was rated by an average of 16 judges. Judges from Studies 1 and 3 each provided ratings for targets' moral characters. I averaged the ratings of Studies 1 and 3 to develop the judges' moral character ratings used in this study.

Results

The descriptive statistics for judges' average ratings of Honesty-Humility, Conscientiousness, and guilt proneness are presented in Table 16. Consistent with the findings of self–judge agreement in Study 4, judges' average rating of Conscientiousness showed the strongest self-judge correlation for the Mistake question. Additionally, consistent with the findings of Study 4, judges' average rating of guilt proneness showed the strongest self–judge correlation for the Dilemma question.

To formally test the relative predictive validity of the judges' average rating of moral character as compared to Honesty-Humility, Conscientiousness, and guilt proneness, I conducted negative binomial regression analyses for each interview question. The results are presented in

Table 17. For the Mistake question, I found that as hypothesized, when both moral character judgments and conscientiousness judgments were entered, the latter was more predictive of unethical behavior. Similarly, for the Dilemma question, I found that guilt proneness judgment had a larger coefficient than that of moral character judgment. When entered together, both ratings were not significant, yet the *p* value of guilt proneness was much smaller than that of moral character judgment.

To formally test the relative predictive validity of self-reports versus judge reports for Honesty-Humility, Conscientiousness, and guilt proneness, I conducted negative binomial regression analyses for each interview question. Consistent with the hypothesis, judges' average rating for Conscientiousness was more predictive of unethical behavior than self-reported Conscientiousness in the Mistake question. Additionally, judges' average rating of guilt proneness was more predictive of unethical behavior than self-reported guilt proneness in the Dilemma question, which supported the hypothesis regarding guilt proneness evaluation.

Discussion

Study 5 provided further evidence that targets' written responses to the Mistake question revealed their Conscientiousness. Consistent with the findings in Study 4, judges' ratings of targets' guilt proneness from the Mistake question were the least informative among all rating dimensions. The lack of validity of judges' guilt proneness evaluation for responses to the mistake question can be explained in two ways. First, it is possible that targets did not talk about their negative emotions at all. Second, it is possible that targets talked about experiencing guilt after making mistakes but that their stated level of guilt after making a mistake was not associated with their actual guilt proneness. The latter explanation is convincing given that guilt proneness is an anticipated experience of bad feelings after wrong-doing. It is possible that

highly guilt-prone individuals are less likely to engage in harmful mistakes that can make them feel bad in the first place and thus are less likely to express those feelings in their responses to the mistake question. I explore this possibility in Chapter 5 by conducting text analyses.

Study 5 found that targets' guilt proneness was better revealed through the dilemma question, which is consistent with the findings in Study 4. Judges' ratings of guilt proneness had stronger predictive power of workplace deviance than overall moral character judgments. Initially, the dilemma question was developed to give targets an opportunity to talk about how their decisions might affect others. Therefore, I expected that the dilemma question would reveal some aspects of targets' Honesty–Humility characteristic. For example, a person who is not narcissistic (i.e., high in modesty) and generous to others (i.e., high in greed avoidance) might be expected to talk about how his or her decisions influence others rather than focusing on himself or herself. However, results from both Studies 4 and 5 suggested that the dilemma question was not good at revealing targets' levels of Honesty-Humility. It is possible that (a) targets did not talk about others in this question or that (b) regardless of whether the targets talked about others in answering the question, this query was not predictive of their unethical behavior. These possibilities are explored in Chapter 5 via text analyses.

Study 6

In Chapter 1, I noted that based on the balanced theory mechanism (Heider, 1957; Insko, 1981), the HIDE model predicts that moral character evaluations from others well-acquainted with the targets may be susceptible to conscious or unconscious bias in evaluations compared to ratings provided by judges who do not know the targets. In Study 6, I tested this theoretical position by comparing the predictive validity of (unacquainted) judges' ratings with those of peer-provided reports on targets' Honesty-Humility, Conscientiousness, and guilt proneness.

Based on the findings of Studies 4 and 5, I hypothesized that judges' average rating of Conscientiousness is more predictive of targets' unethical behavior than the conscientiousness ratings provided by the targets' peers. In Study 6, I used peer-reported CWB as a criterion. Although peer-reported CWB is more favorable to peer-reported independent variables because of the shared method variance, this does not matter because this method decreases type 2 error (i.e., power) and does not increase type 1 error in testing whether judge-reports are more predictive of CWB.

Method

Data Collection from Targets

The target participants in Study 6 were 174 full-time adult U.S. employees recruited from the online participant pool maintained by the university research center.[6] Respondents answered one of three interview questions (Mistake, Dilemma, Employer) and completed the HEXACO-60 personality inventory (Ashton & Lee, 2009) and GP-5 (Cohen, Kim, & Panter, 2014) via a computerized survey. Participants who completed the study were invited via email to to participate in a follow-up study, in which they were asked to invite coworkers to take surveys about them. In total, 87 coworkers participated in the study and provided reports on targets' CWB and completed observer reports of the HEXACO-60 personality inventory.

Data Collection from Judges

Six undergraduate research assistants were recruited to read and evaluate targets' written responses. The order in which targets' responses to the Mistake and Dilemma questions were evaluated by the six judges was randomized. However, responses to the Employer question were the last answers evaluated by the judges, and only four judges completed ratings. In this study,

[6] These participants were a subset of the participants included in a larger project with a larger number of interview questions. The results for the other questions are available from the author.

therefore, I focused on testing theoretical predictions for judges' ratings of the responses to the Mistake and the Dilemma questions. The rating instructions and definitions for on Honesty-Humility, Conscientiousness, guilt proneness, and Agreeableness were the same ones used in Studies 4 and 5.

Results

The descriptive statistics for the self-reported, peer-reported, and judge-rated answers to Honesty-Humility, Conscientiousness, guilt proneness, and Agreeableness questions are presented in Tables 19, 21, and 22. The self–judge correlations and judge–peer correlations for these traits were consistent with the findings in Studies 4 and 5. For the mistake question, judges' Conscientiousness determinations had the strongest convergent validity among judges' ratings. For the Dilemma question, judges' guilt proneness ratings had the strongest convergent validity among judges' ratings.

I formally tested the relative predictive validity of peer- and judge-reported Humility, Conscientiousness, and guilt proneness by conducting negative binomial regression analyses (see Table 23). Consistent with the hypotheses, judges' evaluations of conscientiousness had a stronger predictive power of targets' frequency in engaging in workplace deviance compared to peer-reported Conscientiousness in the Mistake question. Further, the predictive power of judge-reported guilt proneness was stronger than the ratings provided by targets' peers for the Dilemma question condition.

Discussion

In study 2, I demonstrated the validity of judges' moral character ratings based on the criterion of self-reported workplace deviance. In Study 3, I replicated the predictive validity of the judges' evaluations from the text-based interview method with peer-reported CWB.

Consistent with the prediction from the HIDE model, Study 5 found that ratings based on written interview responses provided by judges who did not know the targets were more predictive of targets' CWB than reports from peers well-acquainted with the targets'.

Given that the smaller number of judges providing ratings for targets' written interview responses to the employer question and the order of judgments on the employer question were not randomized, I did not conduct hypothesis testing for judges' evaluations of responses to the Employer question. Nonetheless, I explored the predictive validity of judges' reports for the Employer question and found that only the Conscientiousness evaluation was predictive. Further, when the Conscientiousness evaluations from peers and judges were entered at the same time in the prediction model, only judge-reported Conscientiousness was predictive.

General Discussion

Studies 1, 2, 4, and 5 tested the relative predictive power of self- versus judge-reported moral character ratings. The results of these studies supported the HIDE model prediction that predictive validity of unethical behavior via moral character evaluation would be stronger for judges' reports based on targets' written interview responses than for the ratings directly provided by targets. In Study 6, I further tested the HIDE model prediction that the validity of judges' evaluations of Conscientiousness and guilt proneness based on targets' written responses to the Employer question was mixed across the studies. On the one hand, the results in Study 4 suggested that the employer question revealed targets' guilt proneness but not Conscientiousness. On the other hand, the results in Study 5 suggested that the Employer question might revealing targets' Conscientiousness. In Chapter 4, I further investigate what kind of information is revealed regarding targets' moral characters in their written responses to the Employer question.

CHAPTER IV

Disentangling the Effects of Hiding- and Hidden-Self in the HIDE Model

In the HIDE model, I argue that while ratings directly provided by targets themselves are likely to be influenced by hidden- or hiding-self components, judges' evaluations of targets' written responses to specially designed interview questions can correctly identify information located in those hidden- or hiding-self components. In this chapter, I test this HIDE model prediction by investigating whether hidden- and hiding-self components in self reports can be uncovered from judges' evaluations. If the promise of the HIDE model is true, I expect that the predictive power of judges' evaluations does not decrease when targets' levels of impression management increase, because judges' evaluations are largely based on implicit aspects of targets' moral characters. Moreover, the model predicts that judges' evaluations are less susceptible to the hiding-self components than direct rating provided by targets themselves, because judges are also able to correctly identify implicit aspects of targets' moral characters located in incorrectly-identified-self components in the hiding-self model zone (i.e., impression-management components). I investigate these predictions in studies 7 and 8.

Moreover, Chapter 4 investigates the possible reason of low interpersonal reliability and validity in judges' Honesty-Humility evaluations. In the self-reported personality literature, it is well-established that greed-avoidance is an important element of an individuals' Honesty-Humility, a moral character trait measured in the HEXACO framework. However, I posit that judges' evaluations of greed, which would be perceived as an indicator of unethicality in self reports, is not necessarily interpreted as unethical by judges, but can be interpreted somewhat positively as an indicator of agency and achievement. Because one of the sub-components of Honesty-Humility does not map onto the overarching general factor (i.e., Honesty-Humility), I reason that judges' Honesty-Humility evaluations on the whole are not reliable or valid.

Although it is absolutely true that being extremely greedy can be a negative indicator of moral character, moderate levels of self-promotion might be perceived as moral because they indicate targets' agency and achievement focus. Contemporary moral psychology argues that one important function of morality is to facilitate interpersonal relationships (Cohen et al., 2014; Janoff-Bulman & Carnes, 2013). Combined with the moral character assessment of others, this theory means that people need to evaluate whether targets are going to conduct helpful actions, which can in turn relate to communion. I argue that people also evaluate whether targets are able to conduct those helpful actions, which is closely related to ability aspects of moral character (Cohen & Morse, 201) and ability components in the interpersonal trust model (Mayer, Davis, & Schoorman, 1995). The argument about whether the greed component could be perceived positively by others is also somewhat related to the findings of Walker and Firmer (2007). They examined 50 awardees for either exceptional bravery or caring compared to 50 people in a control group. They found that brave and caring moral exemplars had stronger motivations of both agency and communion than people in the control group. Walker and Firmer's (2007) study is based on self-reported motivation regarding agency and communion, and I investigate whether this can be extended to other reports in this chapter. The ability aspects of moral character, ability components in the interpersonal trust model, and agency component in Walker and Firmer's (2007) study are captured by Conscientiousness in the current study. Conscientiousness is indicative of being "dependable, achievement-striving, hardworking, persevering, and orderly" (Sackett & Walmsley, 2014), which corresponds to ability, dependability, and agency in previous studies defining moral character. Therefore, I investigate whether judges' greed evaluations are positively associated with judges' Conscientiousness evaluations, and are also positively associated with moral character evaluation.

Study 7

In Study 7, I investigate whether the predictive power of judges' moral character evaluations remains still when targets employ different levels of impression management in answering interview questions. Moreover, I investigate whether some portions of targets' impression management components can be revealed by judges' moral character evaluations by testing the interactive effect of impression management and judges' moral character evaluations in predicting targets' unethical behavior.

Data Collection from Targets

The targets participants in this study were 606 U.S. full-time employees recruited from an online participant pool. These participants were assigned to one of the experiment conditions, which were differentiated by the presence of motivation to fake for a reward. Targets in both conditions answered one of the Mistake, the Dilemma, and the Employer questions. Before writing their responses, targets in the reward condition read the following instructions.

> *Imagine that you have been selected to interview for your dream job. The employers want to conduct an online interview before you meet them face-to-face. You will be asked questions about yourself and past experiences you may have had. Please use real examples from your life when responding. Please do not include last names or any other personally identifiable information in your response.*
>
> *When responding to the interview questions and the survey that follows, we would like you to answer as if you are actually applying for a job and attempting to present yourself in the best possible way. The goal is to answer the interview questions in a way that you think would make you appear to be a good person with admirable qualities.*
>
> *Your interview responses and your answers to the personality questions will be evaluated by judges in the future (anonymously). The judges will determine the best job candidates among the participants in this study, based on these responses. Participants who score in the top 5% of the judges' evaluations will be sent a $25 Amazon gift card in a few weeks.*

The participants in the control condition read the instruction that their interview responses would be read by judges in the future. After answering interview questions, the

participants in both conditions completed two questionnaires: the HEXACO-60 personality inventory (Ashton & Lee, 2009) and the GP 5 (Cohen et al., 2014). Before answering the personality questionnaires, the participants in the reward condition were reminded that they should answer the personality questionnaires as if they were actually applying for a job and attempting to present themselves in the best possible way.

To measure to what extent the target participants employed impression management when answering interview questions and personality questionnaires, two questions were administered: In responding to the written interview question, to what extent did you try to answer in a way that would make you appear to be a good person with admirable qualities? *[1 (Not at all), 2 (Slightly), 3 (Moderately), 4 (Quite a bit), 5 (Extremely)]*; In responding to the personality surveys, to what extent did you try to answer in a way that would make you appear to be a good person with admirable qualities? *[1 (Not at all), 2 (Slightly), 3 (Moderately), 4 (Quite a bit), 5 (Extremely)]*.

Finally, participants completed two online tasks (the number task and the problem-solving task) for bonus payments in a randomized order. The number task was based on methods used by Gneezy (2005). In this task, participants were led to believe that they were assigned to one of two possible roles (sender or receiver) and were paired with another participant who played the other role. In reality, all participants were assigned to the sender role. As the sender, participants needed to decide whether to send a deceptive message to the receiver to increase their chances of earning a bonus payment. After participants were given instructions, they completed a comprehension-check test. If they failed the comprehension-check test, they were given the instructions again. If they failed the comprehension check again, they were informed that they could not participate in the number task. The problem-solving task used in Study 1 was

modified to be administered online. Participants were shown a matrix for 7 seconds to find two numbers that add up to 10. Each correctly identified pair of numbers was worth $0.25 in earnings, for a maximum bonus payment of $1.25. Participants indicated whether they solved the matrix once 7 seconds passed. In reality, all matrices were unsolvable, and thus participants who reported that they solved the matrix were considered to have cheated.

Data Collection from Judges

In Study 7, 550 participants were recruited to play the role of judges. These judges were randomly assigned to read written responses to one of three interview question conditions. They were given the same rating instructions and definitions of moral character used in Study 1. Each judge rated the interview responses of 20 randomly selected targets.

Results

Two manipulation-check questions were highly correlated ($r = .70$, $p < .001$), and thus averaged to represent targets' levels of impression management employed in answering questions. The mean of this average score was 2.59 (SD = 1.17) in the control condition and 3.29 (SD = 1.25) in the reward condition. The mean difference between the control condition and reward condition was significant. However, given significant within-group variance, I used the continuous score of the levels of impression management (i.e., the average score itself) rather than using the dummy variable of reward condition.

The frequency-of-lying variable in the target data had missing values because a number of targets (1.5%) failed to pass the comprehension checks in the number task. Because of the existence of these missing values, rather than using the summed count score of lying and cheating, unethical behavior was operationalized by the average score of targets' lying and cheating.

It was found that targets' frequency of cheating and lying did not exactly follow negative binomial or poison distribution. This was because a decent number of targets were concentrated at zero and five, which means that censoring occurred. Consequently, the average frequency of cheating and lying did not exactly follow the normal distribution either. Therefore, to deal with the left- and right-censoring at the same time, when testing hypotheses, I conducted two-sided censored regression analyses.

For each question condition, three different models were analyzed (see Table 27). In the first model, only judges' moral character evaluation was entered. In the second model, targets' level of impression management was also entered. In the third model, the interaction term of targets' level of impression management and judges' moral character evaluation was additionally entered. The results are presented in Table 28. The results indicated that the predictive power of judges' evaluation is strongest in the Mistake question, consistent with the findings in previous studies. In the Mistake question condition, although judges' moral character evaluation was not statistically significant, the coefficient was negative, and its magnitude did not change when the interaction term was entered. However, judges' moral character evaluation did not have predictive power in the Dilemma question condition. For the Employer question, although the moral character judgment did not significantly predict targets' average frequency of cheating and lying, the interaction of judges' evaluation and impression management had a negative prediction, although it was not significant.

Discussion

Consistent with the findings in previous studies, the Mistake question condition revealed the strongest predictive power for judges' moral character evaluation in Study 7. Interestingly and importantly, the interaction term of judges' moral character evaluation and impression

management was negative, which suggests that the predictive power of judges' evaluation actually increased when targets engaged in stronger levels of impression management.

The HIDE model predicts that judges can detect implicit aspects of targets' moral character that targets themselves are unaware of and thus less able to control. It is possible that when targets engaged in impression management when answering the interview question, they actually revealed more of their implicit aspects of moral character than they were able to control. If this prediction is true, it is expected that the predictive power of judges' evaluation should increase as a function of targets' impression management even when controlling targets' self-reported moral character, which is also influenced by their impression management. In Study 8, therefore, I investigated whether the targets' impression management actually increased the predictive power of judges' evaluation.

In this study, judges' evaluation in the Dilemma question did not have any predictive validity. In Chapter 2, I found that the Dilemma question is good to reveal targets' guilt proneness and that the predictive power of the guilt proneness evaluation is stronger than that of moral character. Therefore, it is possible that judges' guilt-proneness evaluation can predict targets' average frequency of cheating and lying. Similarly, the predictive validity in the Mistake and Employer questions are expected to increase when judges evaluate on Conscientiousness s rather than moral character. These hypotheses were invested in Study 8.

Study 8

In Study 8, I focus on judges' ratings on matching dimensions: Conscientiousness evaluation in the Mistake question and guilt proneness evaluation in the Dilemma question. The information revealed by the Employer question is somewhat mixed. Results from Study 4 seemed to suggest that targets' guilt proneness is revealed by the Employer question, but results

from Study 6 seemed to support the claim that Conscientiousness is revealed by the Employer question. Therefore, in Study 8, I focus on both guilt proneness and Conscientiousness evaluations in the Employer question. I investigate whether the predictive validity of judge ratings on the matching evaluation dimension in each question is more predictive when targets employ stronger levels of impression management.

In Study 8, I also investigated how judges form impressions about targets' Honesty-Humility from their written interview responses. In the self-reported personality literature, it was established that greed avoidance is an important element of Honesty-Humility in an individual. However, I propose that targets' greed-avoidance element from the judges' perspective might not be as positive as other elements in Honesty-Humility. I reason that this is one possible reason that overall, judges' Honesty-Humility judgments are not as predictive as other evaluations because subcomponents of overall Honesty-Humility dimensions do not form one factor from judges' evaluation.

Methods

In total, 2,390 participants recruited online served the role of judges. Judges in Study 8 were randomly assigned to eight evaluation dimensions. They read the definition of the evaluation dimension and then read targets on 5-point rating scale ranging from extremely low to extremely high.

> Please evaluate this respondent's tendency to think, feel, and behave in an ethical manner as compared to a typical job applicant. [*Moral Character*]
>
> Please evaluate this respondent's tendency to be fair, sincere, modest, and avoid greed as compared to a typical job applicant. [*Honesty-Humility*]
>
> Please evaluate this respondent's tendency to be organized, diligent, thorough, and inhibit impulses as compared to a typical job applicant. [*Conscientiousness*]

Please evaluate this respondent's tendency to feel bad about his/her mistakes and wrongdoings even if no one knows about them as compared to a typical job applicant. [*guilt proneness*]

Please evaluate this respondent's tendency to be genuine and truthful in his or her interpersonal relations as compared to a typical job applicant. [*Sincerity*]

Please evaluate this respondent's tendency to be fair and avoid fraud/corruption as compared to a typical job applicant. [*Fairness*]

Please evaluate this respondent's tendency to desire lavish wealth, luxury goods, and signs of high social status as compared to a typical job applicant. [*Greed*]

Please evaluate this respondent's tendency to be modest, humble, and unassuming as compared to a typical job applicant. [*Modesty*]

Results & Discussion

The descriptive statistics for Study 8 judges' average ratings of eight dimensions are presented in Table 28. To formally test whether the predictive validity of judges' moral character evaluation from targets' written responses to interview questions increases as targets' levels of impression management increases, two-sided censored regression analyses were conducted for each interview question condition (see Table 30).

For each question condition, three different models were analyzed to compare the predictive power of self-reports versus judge reports. In each model, targets' level of impression management was controlled. In the first model, self-report and interaction of the self-report and impression management were entered. In the second model, judge reports and interaction of the judge report and impression management were entered. In the third model, both the self- and judge-reported main and interaction effects were modeled.

Consistent with the theoretical predictions, judge-reported Conscientiousness was more predictive of targets' unethical behavior in the Mistake question condition. In addition, judges' Conscientiousness evaluation was more predictive of targets' unethical behavior as their levels of

impression management increased. The same pattern was observed for the interaction effect in the Dilemma and Employer questions, such that judges' ratings were predictive of targets' unethical behavior when impression management increases, even though the main effect itself was not significant.

For targets who employed strong levels of impression management (top 25% in the total sample), only the judges' Conscientiousness and guilt proneness evaluations negatively and significantly predicted targets' average levels of cheating and lying in each matching question condition (See Table 31). I formally tested the relative predictive validity of judges' moral character judgments versus Conscientiousness and guilt proneness evaluations, and the results are presented in Tables 32 and 33. Conscientiousness evaluation was more predictive than moral character judgments in the Mistake and Employer question conditions. Guilt proneness judgments were more predictive of targets' average frequency of cheating and lying in the Dilemma question condition.

The correlations between self-reported traits and judge-reports are presented in Table 29. The levels of agreement between targets and judges were the lowest for the greed-avoidance dimension. In particular, there were no correlations between two rating sources, meaning that judge-reported greed avoidance is very dissimilar to self-reported greed avoidance. The correlations among judges' evaluations on all eight dimensions are presented in Table 34. Across all question conditions, I found that judges' greed evaluations were the most positively and strongly correlated with their Conscientiousness evaluations. Greed evaluation was also positively and significantly correlated with moral character evaluations in the Mistake and Employer question conditions. Importantly, greed evaluations were not negatively correlated with Honesty-Humility judgments across all question conditions.

I also investigated the predictive power of four elements of Honesty-Humility in the Employer question, because the Employer question was specifically designed to reveal targets' Honesty-Humility. However, it is possible that only a subset of these four elements might be valid. The results are presented in Table 35. The two-sided censored regression analyses revealed that only the modesty elements have predictive validity regarding targets' unethical behavior when both the Honesty-Humility and modesty evaluations were entered simultaneously.

General Discussion

Findings in Chapter 4 provide further strong evidence of the validity of moral character judgments using the HIDE model. Even when targets' levels of impression management were extremely high, judges' evaluations were predictive of targets' average frequency of cheating and lying. Moreover, while targets' average frequency of cheating and lying were regressed on both the self and judge reports, only the judge reports were predictive. Consistent with the findings in the previous studies, the Mistake question held the strongest predictive power.

In previous chapters, judges' Honesty-Humility did not show good predictive power compared to the other two dimensions. Chapter 4 explored possible reasons and found two important findings. First, only one element of Honesty-Humility, modesty, was predictive of targets' unethical behavior. Second, in contrast to self-reported personality structure, judges' evaluations of fairness, sincerity, modesty, and greed-avoidance do not share enough similarity to form an overarching general factor of Honesty-Humility. This deficiency is largely because the greed component is not as negatively evaluated from judges' perspectives.

CHAPTER V

Comparing Human Judgments to Machine Algorithms

In Chapter 5, I conducted text analyses to explore how human judges utilize linguistic cues in written responses to form an impression of moral character and how linguistic cues predict the unethical behavior of targets. While multiple judges can be an important means to reduce unreliability, certain aspects of unreliability in human judgments are unresolvable when they are due to basic limitations in cognitive capacity or to widely shared cognitive biases (Hammond et al. 1987). The goal of this final chapter was to explore the linguistic cues that human judges failed to correctly detect or utilize In Chapter 5, I used LIWC (Linguistic Inquiry and Word Count) to categorize linguistic cues and patterns in written interview responses, using predefined, high-level word categories. LICW categorizes word usage into higher-order categories and provides information about how frequently these word categories are used in given texts.

According to the stress-emotion model (Fox & Spector, 2006), negative emotions (e.g., frustration, anger) that arise from stressful situations lead individuals to engage in aggressive or harmful behaviors, including workplace deviance (Fox & Spector, 2006). Targets' implicit tendencies to experience negative emotions, therefore, can be predictive of their unethical behavioral tendencies. The Mistake question asks targets how they felt and behaved after making a mistake, which can be a source of stress. Three negative emotions categorized in LIWC are anger, anxiousness, and sadness. Therefore, I investigate whether targets' negative emotions in their written responses to the Mistake question are predictive of judges' moral character evaluations and targets' unethical behaviors.

The Dilemma question was designed to reveal whether targets consider how their own decisions can influence others or only focus on themselves. Therefore, I investigate whether

personal pronoun usage, especially third person pronoun usage, influences judges' moral character evaluations and predict targets' unethical behavior. In LIWC, social process categories (e.g., words related to friends, female/male references, family) can also capture how much targets talked about other people. Finally, a prosocial dictionary (Frimer et al., 2014) consists of words or word stems that are indicative of content about collective interests and interpersonal harmony, which can also be closely related to judges' moral character information from the Dilemma question. Therefore, I investigate whether third-person pronoun usage, social process words, and prosocial words predict targets' unethical behaviors and judges' moral character evaluations of targets.

Targets' written responses to the Employer question could reveal targets' agency and communion focus. In Chapter 4, I argued that targets' agency and achievement focus can be captured by judges' moral character evaluations. In LIWC, affiliation and achievement categories capture individuals' needs, desires, and motivations. Specifically, affiliation category summarizes word usage in reference to others (e.g., ally, social, friend), and achievement category summarizes word usage in reference to success, failure, and achievement striving (e.g., win, success, better). Targets' prosocial dictionaries can also influence judges' moral character evaluations, considering that affiliation motivation is also closely related to collective interests captured in the prosocial dictionary. Therefore, I explore the predictive power of word categories of affiliation, achievement, and prosocial dictionary in the Employer question condition.

Study 9

In total, I analyzed three target data sets. In Study 1, targets answered one each of the Mistake and Dilemma questions and engaged in a problem-solving exercise in which targets' cheating was measured. In Study 2, targets answered one each of the Mistake, Dilemma, and

Employer questions and reported their CWB. In Study 7, targets answered one each of the Mistake, Dilemma, and Employer questions and participated in two online activities that intend to measure targets' cheating and lying.

I conducted lexicon-based text analyses to examine whether theoretically chosen linguistic cues are predictive of unethical behavior and judges' moral character judgments.

Results & Discussion

The correlations between targets' unethical behavior and judges' moral character judgments and LIWC word categories used in analyzing three interview questions (i.e., third-person pronoun usage, three negative emotions, affiliation and achievement, social process words, and prosocial dictionary) are presented in Table 9.

The analysis results for each question condition are presented in Tables 38, 39, and 40. The text analyses revealed that the targets' negative emotions that were revealed in written responses to the Mistake question—especially anger and sadness—were diagnostic of unethical behavior among the targets. Although the judges' moral character evaluation in the Mistake question was negatively associated with anger, it was not associated with sadness. In the Dilemma question, the text analyses revealed that the targets' third-person pronoun usage was a strong, positive predictor of their moral character rating. However, third-person pronoun usage was a positive predictor of unethical behavior in two studies out of three. Finally, in the Employer question, verbal cues of communism (i.e., affiliation) were negative predictors of unethical behavior, and verbal cues of agency (i.e., achievement) was a positive predictor of unethical behavior. However, the judges' evaluations were not predicted by these verbal cues.

Directions of Future Research

Together, the results of Study 9 open the possibility that certain aspects of verbal cues revealed in targets' written responses are not optimally detected or utilized in judges' evaluations. In future work, I aim to investigate more comprehensive sets of verbal cues more systematically to predict targets' unethical behavior and judges' evaluations using machine learning. It is possible that certain verbal cues are better detected and combined by machine-algorithm. Moreover, machine algorithm can quickly process and analyze the latent semantic meanings of large data corpora. In the future work, I will conduct latent semantic analyses (LSA) to identify topics that are predictive of unethical behavior. LSA, which is conceptually similar to factor analysis, is a form of machine-learning for text data that extracts underlying dimensions (i.e., latent semantic clusters). In LSA, each dimension consists of several different linguistic cues (i.e., several words) that appear together in texts. For example, the use of certain keywords in written interviews (e.g., "others", "concern", "worry", "need", "care", "help", "empathize") could reflect semantic factors that would allow us to identify targets who are considerate of others; conversely, targets prone to engage in unethical behaviors would rarely use those keywords. Machine-learning can be used to detect these patterns in written responses.

SUMMARY & CONCLUSION

Chapter 1 of this dissertation introduces a new theoretical framework, the Hidden Information Distribution and Evaluation (HIDE) model. This model enables us to predict that judges, who do not know the targets of evaluations, are able to detect aspects of moral character that the targets misconstrue and/or are unaware of in themselves. Applying the HIDE model to moral character judgments, I developed character interview questions designed to covertly elicit the unethical tendencies of people through their spontaneous written responses.

In chapter 2, I investigated the wisdom of crowds in forecasting unethical behaviors using the text-based interview method that I developed. In studies 1 and 2, I crowd-sourced large sets of judges online and these judges evaluated targets' moral character from written interview responses. In study 1, the judges' average moral character rating negatively and significantly predicted the extent to which targets cheated on the problem-solving task. The predictive power of the judges' average moral character rating was greater than the self-reported moral character traits were, which is consistent with the HIDE model predictions. In study 2, the judges' average moral character rating negatively and significantly predicted the frequency that targets reported engaging in workplace deviance (e.g., falsification of expense reports, stealing, and interpersonal abuse), and supported the external validity of the text-based interview method proposed in this research. Study 3 extended the findings of studies 1 and 2 by determining the judge size at which the crowd effect occurred when forecasting unethical behavior using the text-based interview method. I found that six judges were enough to reliably estimate the moral character of the targets and predict their unethical behaviors.

Having established the possibility of predictive validity in chapter three, in chapter 3, I focused on the aspects of moral character that are elicited by each interview question to improve the predictive power of this text-based interview method. I investigated whether the predictive

validity of the judges' evaluation increased by matching interview questions and evaluation dimensions. The ultimate goal of the chapter 3 was to improve the predictive power of the judges' evaluations by aligning an evaluation dimension to information available from each interview question. I conducted three studies in which judges evaluated the targets on three distinctively defined moral character traits, Honesty-Humility, Conscientiousness, and guilt-proneness.

Study 4 revealed that the convergent validity of the judges' ratings on these three dimensions depended on the interview questions. In particular, the judges' Conscientiousness evaluation had a good convergent validity in the Mistake question whereas the guilt-proneness evaluation had a good convergent validity in the Dilemma question.

In study 5, when the evaluation dimension and the interview question were matched, the judges' average rating on that specific dimension had a stronger predictive power than the judges' average rating on moral character as a whole did. The judges' Conscientiousness evaluation in the Mistake question had a stronger predictive power in regard to cheating by targets than the moral character evaluation did. Likewise, the judges' guilt-proneness evaluation in the Dilemma question had a stronger predictive power in regard to cheating by targets than the moral character evaluation did.

In study 6, I examined another prediction of the HIDE model where the unacquainted judges' evaluation of targets could be more valid than reports made by targets' peers who were well-acquainted with the targets could be. I investigated the relative predictive power of peer-reports versus that of judge-reports with a criterion of workplace deviance reported by the targets' peers. I found that the judges' evaluations of matching dimensions to specific interview questions resulted in stronger predictive powers of workplace deviance than the peer evaluations on these dimensions did.

Chapter 4 investigates two additional research questions. First, I investigated the robustness of the HIDE model predictions under the situation wherein judges employed different levels of impression management. Study 7 revealed that the predictive power of the judges' evaluations of moral character did not decrease when the targets' levels of impression management increased. This finding supports the HIDE model prediction that judges were able to capture aspects of moral character that the targets are unable to control. Second, I investigated how the judges formed an impression of Honesty-Humility in the targets; in the self-reported personality literature, it was established that greed-avoidance is an important element of Honesty-Humility in an individual. However, I hypothesized that the judges' evaluation of greed wass not necessarily interpreted as unethical but can be interpreted more positively as an indicator of agency and achievement orientation. In study 8, my analyses of the relationships among the judges' evaluations of specific elements of Honesty-Humility (i.e., sincerity, fairness, modesty, and greed avoidance) and Honesty-Humility versus Conscientiousness supported this prediction. Greed-evaluation was more strongly and positively correlated with Conscientiousness than Honesty-Humility were. Moreover, despites that greed-avoidance (opposite of greed) was one sub-component of Honesty-Humility in the self-reports, judges' greed evaluation was positively correlated with the Honesty-Humility evaluation in the Mistake question, and it was not correlated with the Honesty-Humility evaluation in the Dilemma and the Employer question conditions.

In chapter 5, I conducted text analyses of the written interview responses from chapters 2 and 4 to explore how the judges made moral character judgments based on target's written interview responses, and to detect linguistic cues that human judges failed to utilize when forming these impressions. The text analyses revealed that the targets' negative emotions that

were revealed in written responses to the Mistake question—especially anger and sadness—were diagnostic of unethical behavior among the targets. Although the judges' moral character evaluation in the Mistake question was negatively associated with anger, it was not associated with sadness. In the Dilemma question, the text analyses revealed that the targets' third-person pronoun usage was a strong, positive predictor of their moral character rating. However, third-person pronoun usage was a positive predictor of unethical behavior in two studies out of three. Finally, in the Employer question, verbal cues of communism (i.e., affiliation) were negative predictors of unethical behavior, and verbal cues of agency (i.e., achievement) was a positive predictor of unethical behavior. However, the judges' evaluations were not predicted by these verbal cues.

Moral character judgment is probably the most important interpersonal judgment. If we can detect moral character of strangers, it would have important practical applications in selection and promotion contexts within organizations, as well as important theoretical implications for understanding how we come to know individuals, and specifically whether they are likely to behave ethically. The most significant contribution of this research is that it supports the notion that strangers' moral character can be detected using the text-based interview method.

Virtually all managers desire an ethical workforce, yet little evidence-based guidance exists for assessing moral character. This study provides evidence that judges can make reasonably reliable and valid judgments of job candidates' moral character based on short written responses to the interview questions. Researchers could use the character-interview questions developed in this research to facilitate understanding of moral character and moral behavior while practitioners could apply the findings from this research to improve personnel selection, promotion, and admissions procedures in organizations.

References

Asendorpf, J. B., & Ostendorf, F. (1998). Is self-enhancement healthy? Conceptual, psychometric, and empirical analysis. *Journal of Personality and Social Psychology, 74*, 955-966.

Blackman, M. C. (2002). Personality Judgment and the Utility of the Unstructured Employment Interview. *Basic and Applied Social Psychology, 24,* 241-250.

Cohen, T. R., Kim, Y., Jordan, K. P., & Panter, A. T. (2016). Guilt-proneness is a marker of integrity and employment suitability. *Personality and Individual Differences, 92*, 109-112.

Cohen, T. R., & Morse, L. (2014). Moral character: What it is and what it does. *Research in Organizational Behavior, 34*, 43-61.

Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2013). Agreement and similarity in self-other perceptions of moral character. *Journal of Research in Personality, 47*, 816-830.

Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2014). Moral character in the workplace. *Journal of Personality and Social Psychology, 107*, 943-963.

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology, 100*, 947-966

Connelly, B. S., & Hülsheger, U. R. (2012). A narrower scope or a clearer lens for personality? Examining sources of observers' advantages over self-reports for predicting performance. *Journal of personality, 80,* 603-631.

Cronbach, L.J., Nageswari, R., & Gleser, G.C. (1963). Theory of generalizability: A liberation of reliability theory. *The British Journal of Statistical Psychology, 16,* 137-163.

Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus in interpersonal perception: acquaintance and the big five. *Psychological bulletin*, *116,* 245-258.

Kim, Y., & Cohen, T. R. (2015). Moral character and workplace deviance: Recent research and current trends. *Current Opinion in Psychology, 6,* 134-138.

Kim, Y., Cohen, T. R., & Panter, A. T. (2016). Cause or consequence? The reciprocal model of counterproductive work behavior and mistreatment. *Academy of Management Annual Meeting Best Paper Proceedings*.

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of personality, 64*, 311-337.

Fast, L. A., & Funder, D. C. (2010). Personality in social psychology. *Handbook of social psychology*, 668-697.

Fleeson, W., Furr, R. M., Jayawickreme, E., Meindl, P., & Helzer, E. G. (2014). Character: The Prospects for a Personality-Based Perspective on Morality. *Social and Personality Psychology Compass*, *8*, 178-191.

Fernandez-Duque, D., & Schwartz, B. (2015). Common Sense Beliefs about the Central Self, Moral Character, and the Brain. *Frontiers in Psychology, 6*.

Funder, D. C. (2012). Accurate personality judgment. Current Directions in *Psychological Science, 21*, 177-182.

Funder, D. C. (1995). On the accuracy of personality judgment: a realistic approach. *Psychological review*, 102, 652.

Funder, D. C., & Colvin, C. R. (1991). Explorations in behavioral consistency: properties of persons, situations, and behaviors. Journal of personality and social psychology, 60, 773.

Gneezy, U. (2005). Deception: The role of consequences. American Economy Review. 95, 384–394.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of personality and social psychology, 106*, 148-168.

Goodwin, G. P. (2015). Moral character in person perception. *Current Directions in Psychological Science, 24*, 38-44.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of personality and social psychology, 74*, 1337.

Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. *Journal of Research in Personality, 57*, 72-88.

Hoevemeyer, V. A. (2005). *High-Impact Interview Questions: 701 Behavior-Based Questions to Find the Right Person for Every Job* (1 edition ed.). New York: AMACOM: American Management Association.

Huffcutt, A. I., Van Iddekinge, C. H., & Roth, P. L. (2011). Understanding applicant behavior in employment interviews: A theoretical model of interviewee performance. *Human Resource Management Review, 21*, 353-367.

Human, L. J., & Biesanz, J. C. (2011). Through the looking glass clearly: accuracy and assumed similarity in well-adjusted individuals' first impressions. *Journal of personality and social psychology, 100*, 349-364.

Insko, C. A. (1981). Balance theory and phenomenology. *Cognitive responses in persuasion*, 309-338.

Larrick, R. P., Mannes, A. E., & Soll, J. B. (2012). The social psychology of the wisdom of crowds. In J. I. Krueger (Ed.), *Frontiers in social psychology: Social judgment and decision making* (pp. 227–242). New York: Psychology Press.

Lee, K., & Ashton, M. C. (2012). *The H Factor of Personality: Why Some People are Manipulative, Self-Entitled, Materialistic, and Exploitive—And Why It Matters for* Everyone. Waterloo, Canada: Wilfrid Laurier University Press.

Luft, J., & Ingham, H. (1955). *Proceedings of the Western Training Laboratory in Group Development*. The Johari windo, a graphic model of interpersonal awareness.

Mayer, R., Davis, J., & Schoorman, F. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review, 20*, 709-734.

Oh, I. S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: a meta-analysis. *Journal of Applied Psychology, 96*, 762-773.

Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the Prediction of Consequential Outcomes. *Annual Review of Psychology, 57*, 401-421.

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. USA: Oxford University Press.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. *Perspectives on Psychological Science, 2*, 313-345.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science, 2*, 313-345.

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., . . . Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology, 103*, 663-688

Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences, 109*, 15197-15200

Stuewig, J., Tangney, J. P., Kendall, S., Folk, J. B., Meyer, C. R., & Dearing, R. L. (2015). Children's proneness to shame and guilt predict risky and illegal behaviors in young adulthood. *Child Psychiatry & Human Development,46*, 217-227.

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of personality and social psychology, 98*, 281-300.

Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: the accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of personality and social psychology, 95*, 1202.

Table 1. Study 1: Descriptive Statistics of Self-Reported Traits and Correlation with Targets' Cheating

	N	Min	Max	Mean	SD	Correlation with Judge-reported Moral Character	Correlation with Cheating
Cheating	195	0.00	16	1.87	3.06		
Honesty-Humility	195	1.50	5.00	3.40	.63	.15*	-.05
Conscientiousness	195	2.10	5.00	3.60	.55	.26*	-.17*
Guilt Proneness	195	1.00	5.00	3.91	.79	.23*	-.10

* $p<.05$

Table 2. Study 1: Descriptive Statistics of Online Judges' Average-Moral-Character-Rating and Correlations with Targets' Cheating

	Total Target N	Total Judge N	Average Judge N	Word Count Mean	Word Count SD	Min	Max	Mean	SD	Correlation with Cheating
Mistake	99	76	15.35	67.71	37.40	1.62	4.33	3.24	.57	-.41***
Dilemma	96	76	15.83	84.98	55.33	2.33	4.63	3.37	.48	-.24*
Total	152	152	15.59	76.21	47.53	1.62	4.63	3.31	.53	-.28***

*** p<.001, ** p<.01, * p<.05

Table 3. Study 1: Negative Binomial Regression of Targets' Number of Cheating on Online Judges' Average-Moral-Character-Judgments

	Judge-report	Self-report	Judge- and Self-reports
	B(S.E.)	B(S.E.)	B(S.E.)
Mistake Question			
Intercept	3.88(.82)***	3.58(1.30)**	4.38(1.25)***
Number correctly solved	-.10(.04)*	-.10(.04)*	-.09(.04)*
Average-Moral-Character-Rating	-.98(.25)***		-.95(.27)***
Honesty-Humility		.04(.30)	.18(.28)
Conscientiousness		-.64(.37)+	-.36(.34)
Guilt proneness		-.15(.18)	.00(.18)
Dilemma Question			
Intercept	3.87(1.11)***	3.24(1.15)**	4.23(1.30)**
Number correctly solved	-.14(.04)*	-.15(.04)***	-.15(.04)***
Average-Moral-Character-Rating	-.72(.32)***		-.61(.36)+
Honesty-Humility		-.10(.33)	-.02(.33)
Conscientiousness		-.41(.26)	-.25(.27)
Guilt proneness		.03(.27)	.06(.27)

*** p<.001, ** p<.01, * p<.05, +p<.10

Table 4. Study 2: Descriptive Statistics of Self-Reported Traits and Correlation with Targets' Counterproductive Work Behaviors

	N	Min	Max	Mean	SD	Correlation with CWB
CWB	798	0	114	64.5	12.58	-.26**
Honesty-Humility	798	2	5	3.65	.60	-.28**
Conscientiousness	798	2	5	3.92	.56	-.25**
Guilt Proneness	798	1	5	4.31	.78	-.26**

** p<.01

Table 5. Study 2: Descriptive Statistics of Online Judges' Average-Moral-Character-Rating and Correlations with Targets' CWB

	Total Target N	Average Judge N	Word Count Mean	Word Count SD	Min	Max	Mean	SD	Correlation with CWB
Mistake	159	16.79	68.29	40.69	2.27	4.59	3.49	.46	-.18*
Dilemma	168	15.78	75.18	43.89	1.94	4.76	3.48	.51	-.16*
Employer	168	16.20	50.2	20.81	1.79	4.43	3.58	.42	-.23*

** p<.01, * p<.05, +<.10

Table 6. Study 2: Negative Binomial Regression of Targets' frequency of CWB

	Judge-Report	Self-Report	Judge- and Self-Reports
	B(S.E.)	B(S.E.)	B(S.E.)
Mistake			
Intercept	5.15(.87)***	8.52(.95)***	8.82(1.06)***
Average-Moral-Character-Rating	-.98(.24)***		-.14(.22)
Honesty-Humility		-.42(.20)*	-.42(.20)*
Conscientiousness		-1.17(.20)***	-1.14(.20)***
Guilt Proneness		-.19(.16)	-.18(.16)
Dilemma			
Intercept	4.22(.85)***	8.44(.96)***	8.76(1.08)***
Average-Moral-Character-Rating	-.67(.24)**		-.16(.24)
Honesty-Humility		-.40(.22)+	-.39(.22)+
Conscientiousness		-.90(.22)***	-.90(.22)***
Guilt Proneness		-.45(.14)**	-.40(.16)*
Employer			
Intercept	3.82(.83)***	6.70(.85)***	7.46(.96)***
Average-Moral-Character-Rating	-.55(.24)*		-.44(.25)+
Honesty-Humility		-.52(.20)**	-.46(.20)*
Conscientiousness		-.45(.19)*	-.39(.19)*
Guilt Proneness		-.34(.13)*	-.26(.14)+

*** p<.001, ** p<.01, * p<.05, +<.10

Table 7. Study 3: Descriptive Statistics of Six Judges' Average-Moral-Character-Rating and Correlations with Targets' Cheating

	Min	Max	Mean	SD	Correlation with Cheating
Mistake	1.50	4.33	3.25	.58	-.41**
Dilemma	2.00	4.67	3.45	.49	-.28**

*** p<.001, ** p<.01, * p<.05

Table 8. Study 3: Correlations between Judge's Moral Character Judgments and Self-Reported Honesty-Humility, Conscientiousness, and Guilt Proneness

	Cheating	Average-Moral-Character-Rating	HH-Self	C-Self
Average-Moral-Character-Rating	-.29**			
Self-reported Honesty-Humility (HH-Self)	-.05	.19**		
Self-reported Conscientiousness (C-Self)	-.16*	.20**	.24**	
Self-reported Guilt proneness	-.10	.30**	.45**	.22**

** p<.01, * p<.05

Table 9. Study 3: G-theory Results: Reliability of Moral Character Judgment as a Function of the Number of Judges

Judge Sample Size	Mistake Question	Dilemma Question
2	0.60	0.50
3	0.70	0.60
4	0.75	0.67
5	0.79	0.72
6	0.82	0.75
7	0.84	0.78
8	0.86	0.80
9	0.87	0.82
10	0.88	0.83
11	0.89	0.85
12	0.90	0.86
13	0.91	0.87
14	0.91	0.88
15	0.92	0.88

Table 10. Study 3: Negative Binomial Regressions of Targets' Number of Cheating on Moral Character Judgements

Models	Mistake Question		Dilemma Question	
	Estimate (S.E.)	p-value	Estimate (S.E.)	p-value
Average-Moral-Character-Rating	-.93 (.24)	<.001	-.78 (.32)	.02
Judge 1	-.56 (.23)	.02	-.16 (.23)	.47
Judge 2	-.86 (.26)	<.01	-.46 (.35)	.18
Judge 3	-.60 (.15)	<.001	-.38 (.19)	.05
Judge 4	-.39 (.17)	.02	-.42 (.14)	<.01
Judge 5	-.68 (.18)	<.001	-.08 (.25)	.76
Judge 6	-.47 (.19)	.01	-.15 (.22)	.49

Table 11. Study 4: Descriptive Statistics of Self-Reported Traits

	N	Min	Max	Mean	S.D.	H	C	GP
Honesty-Humility (H)	296	2.00	7.00	6.21	.95			
Conscientiousness (C)	296	3.00	7.00	6.00	1.08	.53***		
Guilt Proneness (GP)	296	1.00	5.00	4.25	.78	.47***	.30***	
Agreeableness	296	1.50	7.00	5.67	1.19	.49***	.41***	.40***

*** $p < .001$

Table 12. Study 4: Target Size and Word Count Descriptive Statistics

	Total Target N	Total Judge N	Word Count Mean	Word Count SD
Mistake	96	5	68.49	41.77
Dilemma	96	5	79.00	51.48
Employer	44	5	34.66	19.07

Table 13. Study 4: Descriptive Statistics of Judge-Rating-Average

	Min	Max	Mean	SD	H	C	GP
Mistake							
Honesty-Humility (H)	2.60	4.60	3.40	.39			
Conscientiousness (C)	2.20	4.80	3.23	.51	.56***		
Guilt Proneness (GP)	2.20	4.60	3.45	.52	.54***	.53***	
Agreeableness	1.80	4.20	3.25	.38	.57***	.59***	.49***
Dilemma							
Honesty-Humility	2.20	4.60	3.44	.51			
Conscientiousness (C)	1.80	4.60	3.51	.56	.70***		
Guilt Proneness (GP)	2.00	4.40	3.28	.52.	.77***	.73***	
Agreeableness	1.40	4.60	3.30	.52	.26*	.23**	.36**
Employer							
Honesty-Humility	2.00	4.00	3.16	.31			
Conscientiousness (C)	1.60	4.60	3.84	.52	.39*		
Guilt Proneness (GP)	1.60	4.00	3.13	.35	.63***	.70***	
Agreeableness	2.00	4.60	3.40	.58	.37*	.47**	.60***

*** $p < .001$, ** $p < .01$, * $p < .05$

Table 14. Study 4: Reliability of Judge-Rating-Average as a Function of the Number of Judges

Judge Sample Size	Mistake				Dilemma				Employer			
	HH	C	GP	A	HH	C	GP	A	HH	C	GP	A
2	.30	.37	.48	.38	.50	.55	.56	.55	.19	.64	.48	.65
3	.40	.47	.58	.48	.60	.65	.66	.65	.26	.73	.58	.74
4	.47	.54	.65	.55	.66	.71	.72	.71	.32	.78	.65	.79
5	.52	.59	.70	.60	.71	.76	.76	.76	.37	.82	.70	.83
6	.57	.64	.73	.64	.75	.79	.79	.79	.42	.84	.74	.85
7	.60	.67	.76	.68	.78	.81	.82	.81	.46	.86	.77	.87
8	.64	.70	.79	.71	.80	.83	.84	.83	.49	.88	.79	.88
9	.66	.72	.81	.73	.82	.85	.85	.85	.52	.89	.81	.89
10	.69	.74	.82	.75	.83	.86	.86	.86	.55	.90	.82	.90
11	.71	.76	.83	.77	.85	.87	.87	.87	.57	.91	.84	.91
12	.72	.78	.85	.78	.86	.88	.88	.88	.59	.91	.85	.92
13	.74	.79	.86	.80	.87	.89	.89	.89	.61	.92	.86	.92
14	.75	.80	.87	.81	.87	.90	.90	.90	.63	.93	.87	.93
15	.77	.81	.87	.82	.88	.90	.90	.90	.64	.93	.88	.93

Table 15. Study 4: Correlations of Self-Reports and Judge-Reports

	Honesty-Humility	Conscientiousness	Guilt Proneness	Agreeableness
Mistake	.24*	.32**	.01	.16
Dilemmas	.16	.14	.30**	.05
Employer	-.12	-.16	.43**	.12

**:p<.01, *:p<.05

Table 16. Study 5: Descriptive Statistics of Judge-Rating-Average

	Average-Judge N	Min	Max	Mean	SD	Correlation with Self-Reports	Correlation with Cheating
Mistake							
Honesty-Humility	15.56	1.40	4.29	3.25	.63	.23*	-.43***
Conscientiousness	20.81	1.07	4.42	3.06	.65	.29**	-.51***
Guilt Proneness	16.16	1.32	4.65	3.10	.68	.24*	-.27**
Dilemma							
Honesty-Humility	16.25	2.10	4.44	3.31	.46	.02	-.16
Conscientiousness	16.04	1.30	4.23	3.15	.64	.27*	-.42***
Guilt Proneness	15.83	1.71	4.50	3.32	.53	.30**	-.34***

**: p<.001, **: p<.01

Table 17. Study 5: Negative Binomial Regressions of Targets' Number of Cheating on Global versus Specific Trait Judgments

	Mistake Question		Dilemma Question	
	Estimate (S.E.)	p-value	Estimate (S.E.)	p-value
Moral Character Judgment	-.39 (.68)	.57	-.91 (.47)	.05
Honesty-Humility Judgment	-.60 (.60)	.32	.09 (.45)	.84
Moral Character Judgment	.32 (.50)	.52	.52 (.56)	.36
Conscientiousness Judgment	-1.33 (.44)	.00	-1.09 (.38)	.00
Moral Character Judgment	-1.31 (.42)	.00	-.20 (.61)	.74
Guilt Proneness Judgment	.30 (.34)	.38	-.57 (.47)	.22

Table 18. Study 5: Negative Binomial Regressions of Targets' Number of Cheating on Self- and Judge-reported Honesty-Humility, Conscientiousness, and Guilt Proneness

Models	Mistake Question		Dilemma Question	
	Estimate (S.E.)	p-value	Estimate (S.E.)	p-value
A1. Honesty-Humility (Judge)	-.92 (.22)	.00	-.53 (.34)	.11
A2. Honesty-Humility (Self)	-.28 (.26)	.28	-.24 (.26)	.37
A3. Honesty-Humility (Combined)				
Judge-report	-.96 (.23)	.00	-.48 (.34)	.16
Self-report	.14 (.25)	.59	-.12 (.27)	.65
B1. Conscientiousness (Judge)	-1.09 (.21)	.00	-.81 (.23)	.00
B2. Conscientiousness (Self)	-.68 (.33)	.04	-.44 (.25)	.07
B3. Conscientiousness (Combined)				
Judge-report	-1.07 (.23)	.00	-.77 (.24)	.00
Self-report	-.05 (.30)	.86	-.14 (.24)	.57
C1. Guilt Proneness (Judge)	-.57 (.21)	.01	-.70 (.26)	.01
C2. Guilt Proneness (Self)	-.21 (.17)	.22	-.10 (.22)	.65
C3. Guilt Proneness (Combined)				
Judge-report	-.54 (.22)	.01	-.77 (.28)	.00
Self-report	-.10 (.17)	.57	.16 (.23)	.49

Note. A1, B1, C1: only the judge-reported variable is used as a predictor; A2, B2, C2: only the self-reported variable is used as a predictor; A3, B3, C3: both the judge- and self-reported variables are used as predictors.

Table 19. Studies 6: Descriptive Statistics of Self- and Peer-Reported Traits

	N	Min	Max	Mean	S.D.	H	C	GP
Self-Reports								
Honesty-Humility (H)	171	1.10	4.80	3.31	.64			
Conscientiousness (C)	172	2.00	5.00	3.72	.53	.17*		
Guilt Proneness (GP)	172	1.00	5.00	3.92	.87	.36***	.31***	
Agreeableness	171	1.30	4.60	3.20	.60	.25**	.05	.18*
Peer-Reports								
Honesty-Humility (HH)	87	1.70	4.70	3.37	.52			
Conscientiousness (C)	87	1.80	4.70	3.83	.55	.44***		
Guilt Proneness (GP)	87	1.40	5.00	4.19	.80	.25*	.08	
Agreeableness	87	1.40	4.70	3.33	.61	.47***	.11	.00

**: p<.001, *: p<.01

Table 20. Study 6: Target Size and Word Count Descriptive Statistics

	Total Target N	Total Judge N	Word Count Mean	Word Count SD
Mistake	59	6	61.90	32.95
Dilemma	59	6	76.22	46.19
Employer	56	4	29.89	14.29

Table 21. Study 6: Self-Judge Correlations and Judge-Peer Correlations Across Interview Question Conditions

	Self-Judge Correlations				Judge-Peer Correlations			
	HH	C	GP	A	HH	C	GP	A
Mistake	-.05	.20	.08	-.10	-.18	.28	.13	.15
Dilemmas	.25+	.05	.17	.21	.22	-.05	.48**	.20
Employer	.12	.31*	-.16	.01	-.30	.01	-.32+	-.23

**:p<.01, +: p ≤.10

Table 22. Study 6: Descriptive Statistics of Judges' Rating Average

	Min	Max	Mean	SD	H	C	GP
Mistake							
Honesty-Humility (H)	2.00	4.33	3.40	.40			
Conscientiousness (C)	2.00	4.17	3.23	.37	.39**		
Guilt Proneness (GP)	2.50	4.83	3.52	.53	.36**	.40**	
Agreeableness	2.67	4.00	3.21	.26	.67***	.38**	.49***
Dilemma							
Honesty-Humility	2.00	4.67	3.42	.49			
Conscientiousness (C)	2.17	4.50	3.56	.49	.61***		
Guilt Proneness (GP)	2.17	4.67	3.25	.46	.81***	.72***	
Agreeableness	2.17	4.50	3.37	.47	.56***	.57***	.68***
Employer							
Honesty-Humility	2.25	4.00	3.18	.36			
Conscientiousness (C)	2.25	4.50	3.71	.52	.15		
Guilt Proneness (GP)	2.75	3.50	3.10	.17	.49***	.24+	
Agreeableness	2.75	4.50	3.62	.40	.33*	.28*	.29*

*** $p < .001$, ** $p < .01$, * $p < .05$, +: $p \leq .10$

Table 23. Study 6: Negative Binomial Regressions of Peer-Reported CWB on Peer- and Judge-Reported Traits

	Model 1		Model 2		Model 3	
	B (S.E.)	p-value	B (S.E.)	p-value	B (S.E.)	p-value
Mistake						
Conscientiousness (Judge)	-.23 (.47)	.32			-.35 (.51)	.25
Conscientiousness (Peer)			.14 (.33)	.33	.23 (.35)	.26
Dilemma						
Guilt Proneness (Judge)	-.76 (.69)	.14			-.76 (.69)	.14
Guilt Proneness (Peer)			-.01 (.55)	.49	-.01 (.55)	.49
Employer						
Conscientiousness (Judge)	-1.11 (.62)	.04			-1.11 (.61)	.03
Conscientiousness (Peer)			-.35 (.48)	.23	-.34 (.44)	.22

Table 24. Study 7: Descriptive Statistics of Self-Reported Traits

	N	Mean	S.D.	Conscientiousness	Guilt Proneness
Mistake					
Honesty-Humility	201	3.47	.70		
Conscientiousness	201	3.82	.69	.51***	
Guilt Proneness	201	4.01	.85	.56***	.37***
Dilemma					
Honesty-Humility	195	3.57	.72		
Conscientiousness	195	3.97	.63	.43***	
Guilt Proneness	195	4.05	.79	.55***	.49***
Employer					
Honesty-Humility	210	3.67	.69		
Conscientiousness	210	4.02	.60	.49***	
Guilt Proneness	210	4.18	.78	.57***	.46***

***: $p < .001$, **: $p < .01$, *: $p < .05$

Table 25. Study 7: Descriptive Statistics of Responses to Interview Questions

	Total Target N	Word Count Mean	Word Count SD
Mistake	210	63.06	46.07
Dilemma	201	95.13	57.02
Employer	195	128.16	71.96

Table 26. Study 7: Descriptive Statistics of Judges' Average Moral Character Rating

	Average Judge N	Min	Max	Mean	SD
Mistake	17.61	1.52	4.45	3.34	.54
Dilemma	17.85	2.04	4.64	3.51	.51
Total	16.87	1.88	4.39	3.55	.42

Table 27. Study 7. Two-Sided Censored Regression of Targets' Average Frequency of Cheating and Lying on Judges' Average Moral Character Rating

	Model 1		Model 2		Model 3	
	B (SE)	p-value	B (SE)	p-value	B (SE)	p-value
Mistake						
Intercept	2.44 (.15)	.00	2.44 (.15)	.00	2.45 (.15)	.00
Average-Moral-Character-Rating (MC)	-.12 (.14)	.19	-.13 (.14)	.17	-.12 (.14)	.19
Impression Management (IM)			.32 (.15)	.02	.27 (.16)	.04
IIM × MC					-.16 (.14)	.11
Dilemma						
Intercept	2.86 (.16)	.00	2.86 (.16)	.00	2.86 (.16)	.00
Average-Moral-Character-Rating (MC)	-.00 (.15)	.50	-.01 (.15)	.47	-.01 (.15)	.47
Impression Management (IM)			.18 (.16)	.12	.18 (.16)	.13
IIM × MC					.02 (.15)	.45
Employer						
Intercept	2.29 (.17)	.00	2.30 (.17)	.00	2.33 (.18)	.00
Average-Moral-Character-Rating (MC)	.19 (.20)	.17	.14 (.21)	.25	.08 (.23)	.37
Impression Management (IM)			.20 (.17)	.12	.25 (.19)	.09
IIM × MC					-.16 (.22)	.23

Note. One-tail test is conducted given that the hypothesis is one-sided.

Table 28. Studies 8: Descriptive Statistics of Judges' Average-Ratings across Evaluation Conditions

	Mistake			Dilemma			Employer		
	Average Judge N	Mean	SD	Average Judge N	Mean	SD	Average Judge N	Mean	SD
Global Judgment	7.09	3.18	.55	11.03	3.42	.53	11.70	3.46	.47
Honesty-Humility	10.07	3.26	.58	9.87	3.39	.53	9.94	3.36	.45
Conscientiousness	10.70	2.93	.61	10.38	3.27	.68	8.31	3.43	.63
Guilt Proneness	10.57	3.14	.70	9.62	3.17	.54	9.38	3.02	.43
Fairness	9.08	3.26	.59	10.51	3.38	.63	9.69	3.33	.44
Sincerity	10.70	3.27	.62	9.87	3.38	.55	9.59	3.35	.44
Modesty	10.32	3.15	.44	10.90	3.15	.47	9.86	2.81	.52
Greed Avoidance	10.95	3.16	.63	8.72	3.24	.50	9.11	3.05	.48

Table 29. Study 8: Correlations of Self-Reports and Judge-Reports across Interview Question Conditions

	Mistake	Dilemma	Employer
Honesty-Humility	.24**	.23**	.08
Conscientiousness	.40***	.36***	.28***
Guilt Proneness	.16*	.17*	.11
Fairness	.21**	.23**	.13+
Sincerity	.03	.12+	.13+
Modesty	.17*	.11	.19**
Greed Avoidance	-.00	-.06	.05

*** $p < .001$, ** $p < .01$, * $p < .05$, +: $p \leq .10$

Table 30. Study 8. Two-Sided Censored Regression of Targets' Average Frequency of Cheating and Lying on Self- and Judge-Reported Traits

	Model 1		Model 2		Model 3	
	B (SE)	p-value	B (SE)	p-value	B (SE)	p-value
Mistake						
Intercept	2.43 (.15)	.00	2.34 (.16)	.00	2.44 (.15)	.00
IM	.32 (.15)	.01	.19 (.17)	.12	.30 (.15)	.02
C (Self)	-.27 (.14)	.03			-.17 (.15)	.12
IM × C (Self)	-.13 (.14)	.17			-.01 (.16)	.47
C (Judge)			-.31 (.17)	.03	-.21 (.16)	.10
IM × C (Judge)			-.24 (.16)	.06	-.19 (.17)	.12
Dilemma						
Intercept	2.85 (.15)	.00	2.87 (.16)	.00	2.84 (.15)	.00
IM	.18 (.15)	.12	.17 (.15)	.13	.16 (.16)	.16
GP (Self)	-.32 (.16)	.02			-.30 (.16)	.03
IM × GP (Self)	-.02 (.15)	.46			.02 (.16)	.45
GP (Judge)			-.15 (.16)	.17	-.09 (.16)	.27
IM × GP (Judge)			-.11 (.16)	.24	-.11 (.16)	.24
Employer						
Intercept	2.43 (.15)	.00	2.34 (.16)	.00	2.42 (.18)	.00
IM	.32 (.15)	.01	.19 (.17)	.12	.39 (.19)	.02
C (Self)	-.27 (.14)	.03			-.03 (.19)	.43
IM × C (Self)	-.13 (.14)	.17			-.14 (.20)	.24
C (Judge)			-.31 (.17)	.03	-.15 (.19)	.22
IM × C (Judge)			-.24 (.16)	.06	-.32 (.22)	.07
Employer						
Intercept	2.39 (.17)	.00	2.35 (.17)	.00	2.42 (.17)	.00
IM	.37 (.18)	.02	.16 (.17)	.19	.29 (.18)	.05
GP (Self)	-.56 (.18)	.00			-.56 (.18)	.00
IM × GP (Self)	-.05 (.16)	.38			-.03 (.16)	.42
GP (Judge)			.08 (.22)	.35	.16 (.22)	.24
IM × GP (Judge)			-.40 (.22)	.04	-.37 (.22)	.05

Note. One-tail test is conducted given that the hypothesis is one-sided.

Table 31. Study 8. Two-Sided Censored Regression of Targets' Average Levels of Cheating and Lying on Self- and Judge-Reports under High Levels of Impression Management

	Model 1		Model 2		Model 3	
	B (SE)	p-value	B (SE)	p-value	B (SE)	p-value
Mistake						
Intercept	3.18 (.31)	.00	2.83 (.34)	.00	2.89 (.36)	.00
C (Self)	-.41 (.26)	.06			-.15 (.31)	.31
C (Judge)			-.62 (.30)	.02	-.52 (.37)	.08
Dilemma						
Intercept	3.00 (.34)	.00	2.96 (.33)	.00	2.96 (.33)	.00
GP (Self)	-.33 (.35)	.18			-.05 (.37)	.45
GP (Judge)			-.81 (.36)	.01	-.79 (.39)	.02
Employer						
Intercept	2.25 (.44)	.00	2.58 (.40)	.00	2.37 (.46)	.00
C (Self)	.28 (.46)	.27			.45 (.49)	.18
C (Judge)			-.36 (.45)	.21	-.50 (.48)	.15
Employer						
Intercept	2.45 (.46)	.00	2.36 (.34)	.00	2.39 (.45)	.00
GP (Self)	-.04 (.45)	.47	-.72 (.51)	.08	-.05 (.44)	.45
GP (Judge)					-.72 (.51)	.08

Note. Mistake: Target N=42; Dilemma: Target N=50; Employer: N=47.

Table 32. Study 8: Two-Sided Censored Regressions of Targets' Average Number of Cheating and Lying on Moral Character versus Conscientiousness Judgments

	Mistake Question		Employer Question	
	Estimate (S.E.)	p-value	Estimate (S.E.)	p-value
Intercept	2.35 (.16)	.00	2.36 (.18)	.00
IM	.29 (.15)	.03	.23 (.17)	.09
Moral Character Judgments	.26 (.20)	.09	.21 (.25)	.21
Conscientiousness Judgments	-.56 (.22)	.01	-.25 (.24)	.14

Table 33. Study 8: Two-Sided Censored Regressions of Targets' Average Number of Cheating and Lying on Moral Character versus Guilt Proneness Judgments

	Dilemma Question		Employer Question	
	Estimate (S.E.)	p-value	Estimate (S.E.)	p-value
Intercept	2.88 (.16)	.00	2.35 (.18)	.00
IM	.18 (.16)	.12	.23 (.17)	.10
Moral Character Judgments	-.08 (.17)	.31	-.02 (.22)	.46
Guilt Proneness Judgments	-.11 (.18)	.28	.14 (.25)	.29

Table 34. Study 8. Correlations among Judges' Evaluations on Targets' Moral character, Honesty-Humility, Conscientiousness, Guilt Proneness, and Four Elements of Honesty-Humility

	Moral character	HH	C	GP	Fairness	Sincerity	Modesty
Mistake							
Honesty-Humility (HH)	.81						
Conscientiousness (C)	.73	.67					
Guilt Proneness (GP)	.67	.68	.47				
Fairness	.78	.73	.74	.62			
Sincerity	.72	.67	.67	.65	.72		
Modesty	.76	.74	.55	.74	.68	.65	
Greed	.23**	.26***	.37***	.13+	.28***	.29***	.17*
Dilemma							
Honesty-Humility (HH)	.77						
Conscientiousness (C)	.70	.69					
Guilt Proneness (GP)	.48	.41	.31				
Fairness	.78	.68	.67	.38			
Sincerity	.69	.67	.67	.49	.66		
Modesty	.62	.60	.59	.39	.51	.48	
Greed	.00	.04	.22**	-.13+	.14+	.12+	-.05 (n.s.)
Employer							
Honesty-Humility (HH)	.46						
Conscientiousness (C)	.65	.34					
Guilt Proneness (GP)	.50	.26	.44				
Fairness	.71	.31	.60	.43			
Sincerity	.21	.23	.17	.25	.25		
Modesty	.21	.45	.08	.13	.15	.26	
Greed	.35***	-.05	.43***	.13+	.32***	.01	-.34***

***: p<.001, **: p<.01, *: p<.05

Note. Correlations among HH, C, GP, Fairness, Sincerity and Modesty are all significant with α=.001.

Table 35. Study 8. Two-Sided Censored Regression of Targets' Average Levels of Cheating and Lying on Honesty-Humility Factors and Facets in the Employer Condition

	Fairness		Sincerity		Modesty		Greed Avoidance	
	B (SE)	p-value	B (SE)	p-value	B (SE)	p-value	B (SE)	p-value
Intercept	2.32 (.17)	.00	2.32 (.17)	.00	2.27 (.18)	.00	2.31 (.19)	.00
IM	.23 (.17)	.09	.23 (.17)	.09	.24 (.17)	.03	.23 (.17)	.09
Honesty-Humility	-.02 (.21)	.46	-.01 (.21)	.47	.09 (.22)	.04	-.01 (.20)	.47
Facet	.01 (.23)	.48	-.01 (.22)	.48	-.23 (.22)	.04	-.03 (.17)	.43

Table 36. Study 8. Censored Regression of Targets' Average Levels of Cheating and Lying on Honesty-Humility Sub-components in the Employer Condition under High Levels of Impression Management

	Fairness		Sincerity		Modesty		Greed Avoidance	
	B (SE)	p-value	B (SE)	p-value	B (SE)	p-value	B (SE)	B (SE)
Intercept	2.41 (.35)	.00	2.44 (.36)	.00	2.42 (.34)	.00	2.06 (.38)	.00
Average-Judge-Rating	.10 (.45)	.41	.07 (.44)	.44	-.36 (.43)	.20	-.75 (.39)	.03

Table 37. Study 9: Text Cue Correlations with Targets' Unethical Behavior and Judges' Moral Character Evaluation

	Unethical Behavior			Moral Character Judgments		
	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3
			Mistake			
Pronouns						
First Person Pronoun Ratio	-.16+	-.19**	.02	-.03	-.02	-.06
Third Person Pronoun Ratio	.34***	.04	-.08	.02	.05	.05
Negative Emotions						
Anger	.43***	.01	.14*	-.33**	-.15	-.16*
Anxious	-.18*	.10	.11+	.17+	.09	-.03
Sad	.35***	.20**	.07	-.04	-.03	-.04
Agency vs Communion						
Affiliation	-.10	.02	.01	.01	.14*	.03
Achievement	.10	.08	.01	.03	-.06	.03
Social Words						
Prosocial Words	-.24*	.05	-.10+	.18*	.12+	.11+
Social Words	.09	.11+	.06	-.06	0.04	-.03
			Dilemma			
Pronouns						
First Person Pronoun Ratio	-.12	-.08	.04	-.29**	-.20**	-.19**
Third Person Pronoun Ratio	.13	.19**	-.19**	.31**	.19**	.10+
Negative Emotions						
Anger	-.18	.03	-.02	.18*	.10	.02
Anxious	.17	.16*	.09	-.10	-.02	-.05
Sad	-.16	.02	-.04	-.11	-.05	.02
Agency vs Communion						
Affiliation	-.09	-.07	.10+	.24*	.07	.05
Achievement	.01	.08	.06	.16+	-.11+	.06
Social Words						
Prosocial Words	.10	-.08	-.07	.10	.08	.08
Social Words	-.11	-.16*	-.08	.26**	.22**	.09+
			Employer			
Pronouns						
First Person Pronoun Ratio	.03	.00	-.21**	.11+	.03	.00
Third Person Pronoun Ratio	.00	-.02	.18*	.01	.00	-.02
Negative Emotions						
Anger	-.02	.01	-.02	-.18*	-.02	.01
Anxious	-.07	.04	.06	.06	-.07	.04
Sad	.01	-.06	.01	.01	.01	-.06
Agency vs Communion						
Affiliation	-.19**	-.02	.15*	.28***	-.19**	-.02
Achievement	-.12+	.03	.07	-.06	-.12+	.03
Social Words						
Prosocial Words	-.09	-.10+	.31***	.24***	-.09	-.10+
Social Words	.02	-.03	0.09	.11+	.02	-.03

*** $p < .001$, ** $p < .01$, * $p < .05$, +: $p \leq .10$

Table 38. Study 9 (Mistake Question): Text Cue Predictions of Targets' Unethical Behavior on Judges' Moral Character Judgment

	B (SE)	p-value	B (SE)	p-value	B (SE	p-value	B (SE)	p-value
Prediction of Unethical Behavior								
Sample 1								
Number Correctly Solved	-.13 (.04)	.00	-.10 (.04)	.01	-.09 (.04)	.02	-.14 (.04)	.00
Anger	.28 (.11)	.00					.27 (.11)	.01
Anxiousness			-.02 (.12)	.45			-.36 (.18)	.03
Sadness					.23 (.13)	.03	.10 (.13)	.26
Sample 2								
Anger	.02 (.17)	.46					.17 (.18)	.17
Anxiousness			.15 (.11)	.08			.23 (.11)	.02
Sadness					.28 (.14)	.02	.35 (.14)	.01
Sample 3								
IM	.28 (.15)	.03	.31 (.15)	.02	.14 (.15)	.17	.28 (.18)	.06
Anger	.39 (.17)	.01					-.31 (.53)	.28
IM × Anger	-.21 (.16)	.09					-.38 (.28)	.09
Anxiousness			.21 (.15)	.08			.25 (.45)	.29
IM × Anxiousness			.12 (.15)	.21			-1.86 (.79)	.01
Sadness					.32 (.15)	.02	.43 (.45)	.17
IM× Sadness					.01 (.14)	.49	.07 (.38)	.43
Prediction of Moral Character Judgment								
Sample 1								
Anger	-.15 (.04)	.00					-.15 (.05)	.00
Anxiousness			.08 (.05)	.05			.08 (.05)	.05
Sadness					-.02 (.05)	.35	.04 (.05)	.23
Sample 2								
Anger	-.09 (.05)	.03					-.09 (.05)	.03
Anxiousness			.04 (.03)	.13			.03 (.03)	.16
Sadness					-.02 (.04)	.35	-.02 (.04)	.34
Sample 3								
Anger	-.09 (.04)	.01					-.10 (.04)	.01
Anxiousness			-.01 (.04)	.35			-.02 (.04)	.32
Sadness					-.03 (.05)	.28	-.04 (.05)	.20

Table 39. Study 9 (Dilemma Question): Text Cue Predictions of Targets' Unethical Behavior and Moral Character Judgments

	B (SE)	p-value	B (SE)	p-value	B (SE)	p-value	B (SE)	p-value
Prediction of Unethical Behavior								
Sample 1								
Number Correctly Solved	-.14 (.04)	.00	-.15 (.04)	.00	-.38 (.09)	.00	-.17 (.04)	.00
Third Person Pronoun Ratio	.57 (.81)	.24					-.18 (1.02)	43
Social Words			.03 (.03)	.13			.01 (.04)	.43
Prosocial Dictionary					.25 (.14)	.04	.10 (.07)	.08
Sample 2								
Third Person Pronoun Ratio	.99 (.48)	.02					1.83 (.60)	.00
Social Words			-.06 (.02)	.00			-.09 (.03)	.00
Prosocial Dictionary					-.12 (.08)	.07	-.06 (.08)	.22
Sample 3								
IM	.14 (.15)	.18	.17 (.15)	.14	.19 (.15)	.12	.14 (.15)	.18
Third Person Pronoun Ratio	-.49 (.15)	.00					-.51 (.20)	.00
IM × Third.	-.32 (.16)	.02					-.36 (.20)	.04
Social Words			-.21 (.16)	.09			.10 (.20)	.30
IM × Social Words			-.06 (.15)	.34			.13 (.19)	.24
Prosocial Dictionary					-.16 (.16)	.14	-.10 (.16)	.26
IM × Prosocial Dic.					-.17 (.15)	.12	-.12 (.16)	.22
Prediction of Moral Character Judgment								
Sample 1								
Third Person Pronoun Ratio	.72 (.23)	.00					.55 (.27)	.02
Social Words			.02 (.01)	.00			.01 (.01)	.15
Prosocial Dictionary					.02 (.02)	.16	.00 (.02)	.49
Sample 2								
Third Person Pronoun Ratio	.45 (.18)	.00					.27 (.20)	.10
Social Words			.02 (.01)	.00			.02 (.01)	.05
Prosocial Dictionary					.02 (.02)	.14	.00 (.02)	.42
Sample 3								
Third Person Pronoun Ratio	.32 (.20)	.06					.27 (.25)	.15
Social Words			.01 (.01)	.10			.00 (.01)	.43
Prosocial Dictionary					.03 (.02)	.13	.02 (.03)	.18

*** p < .001, ** p < .01, * p < .05, +: p ≤.10

Table 40. Study 9 (Employer Question): Text Cue Predictions of Targets' Unethical Behavior

	B (SE)	p-value	B (SE)	p-value	B (SE)	p-value	B (SE)	p-value
Prediction of Unethical Behavior								
Sample 2								
Affiliation	-.16 (.04)	.00					-.15 (.05)	.00
Achievement			-.07 (.03)	.01			-.04 (.03)	.03
Prosocial Dictionary					-.08 (.04)	.02	.00 (.04)	.48
Sample 3								
IIM	.23 (.17)	.23 (.17)	.27 (.17)	.06	.24 (.17)	.08	.30 (.17)	.04
Affiliation	-.05 (.17)	-.05 (.17)					.06 (.19)	.38
IM × Affiliation	.00 (.16)	.00 (.16)					-.04 (.17)	.39
Achievement			.13 (.17)	.23			.10 (.17)	.28
IM × Achievement			.39 (.18)	.01			.44 (.19)	.01
Prosocial Dictionary					-.26 (.17)	.07	-.31 (.19)	.05
IM × Prosocial. Dic.					.04 (.19)	.43	.16 (.20)	.21
Prediction of Moral Character Judgment								
Sample 2								
Affiliation	.02 (.01)	.02					.00 (.07)	.47
Achievement			.01 (.01)	.31			.01 (.01)	.22
Prosocial Dictionary					.04 (.01)	.00	.04 (.01)	.00
Sample 3								
Affiliation	.05 (.01)	.00					.04 (.01)	.00
Achievement			-.01 (.01)	.19			.00 (.01)	.21
Prosocial Dictionary					.03 (.01)	.00	.02 (.01)	.03

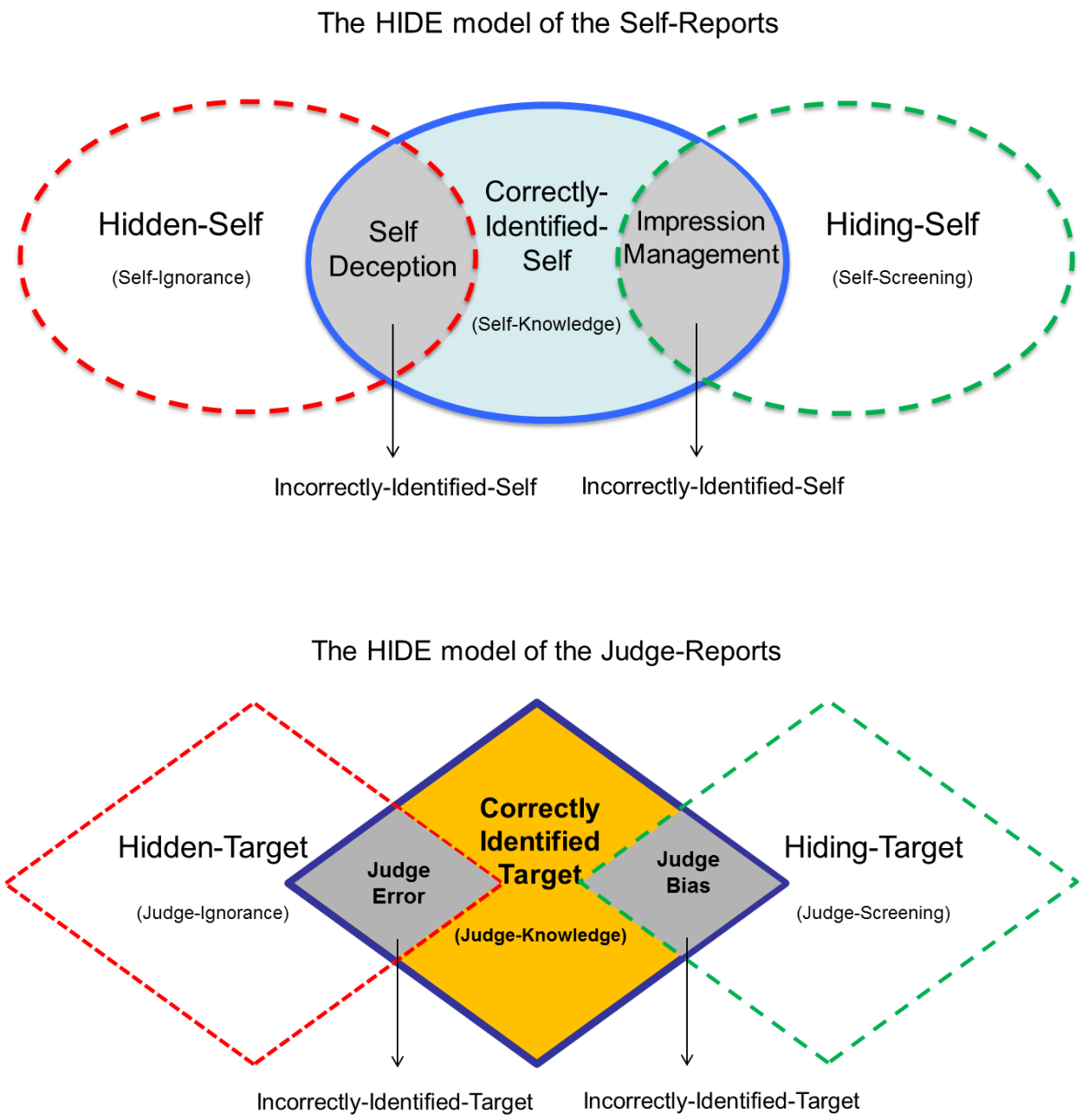

Figure 1. The Hidden Information Distribution and Evaluation (HIDE) Model of Person Perception

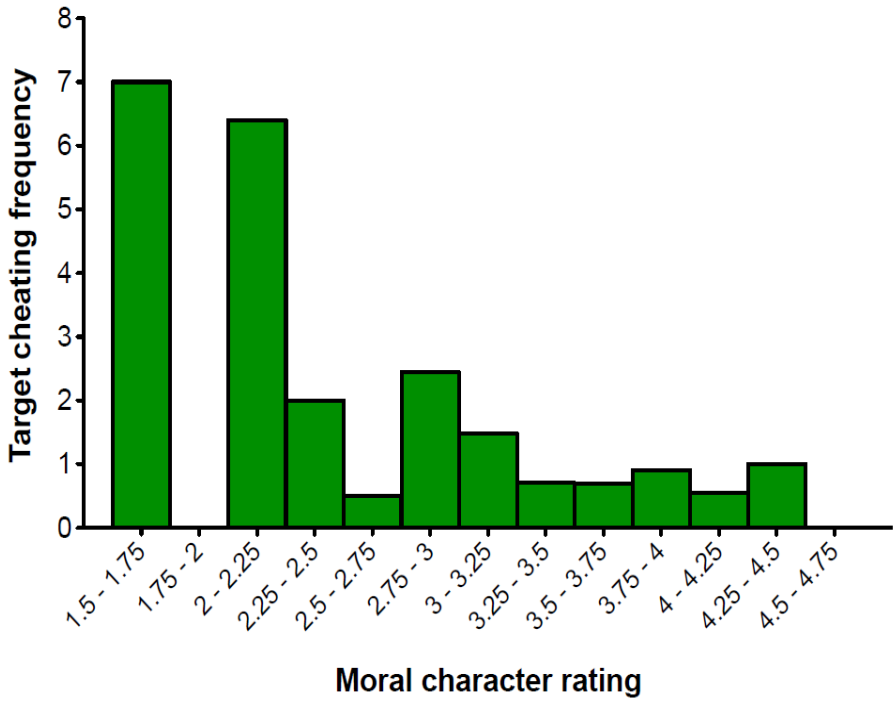

Figure 2. Study 1: Cheating in the Problem-Solving Task as a function of Judges' Ratings of Moral Character for Mistake Question

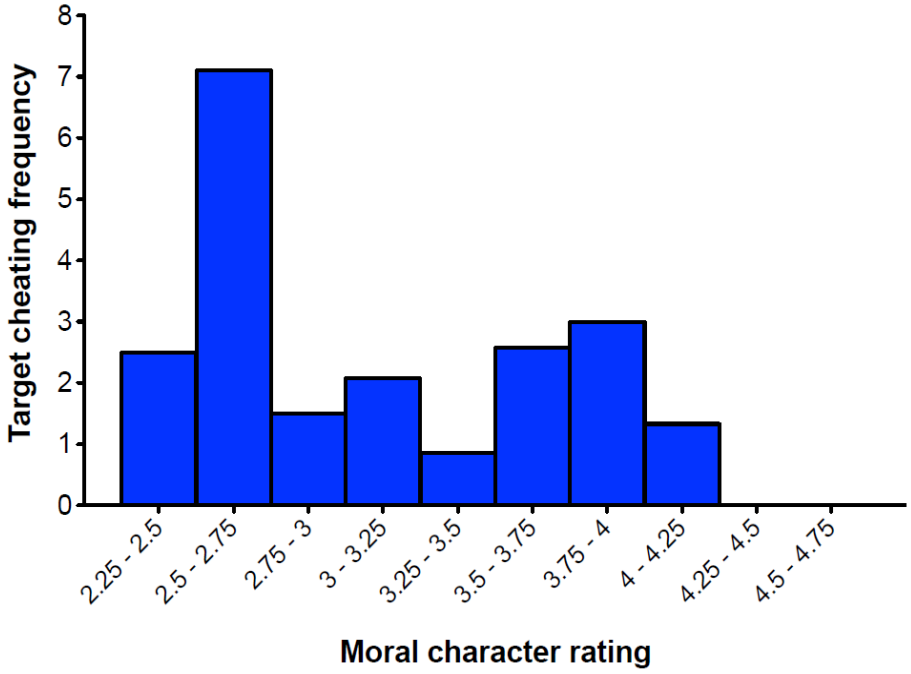

Figure 3. Study 1: Cheating in the Problem-Solving Task as a function of Judges' Ratings of Moral Character for Dilemma Question

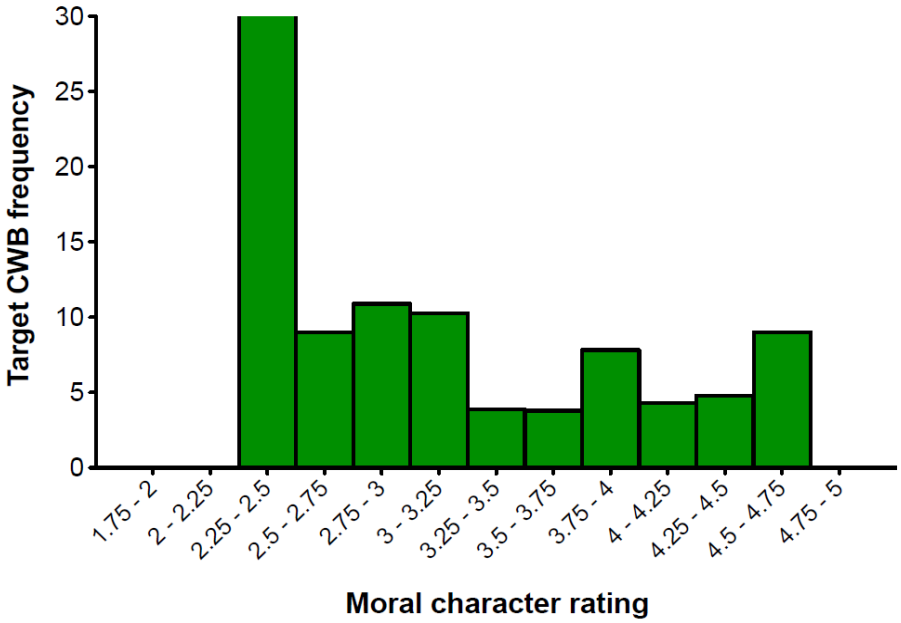

Figure 2. Study 2: Frequency of CWB as a function of Judges' Ratings of Moral Character for Mistake Question

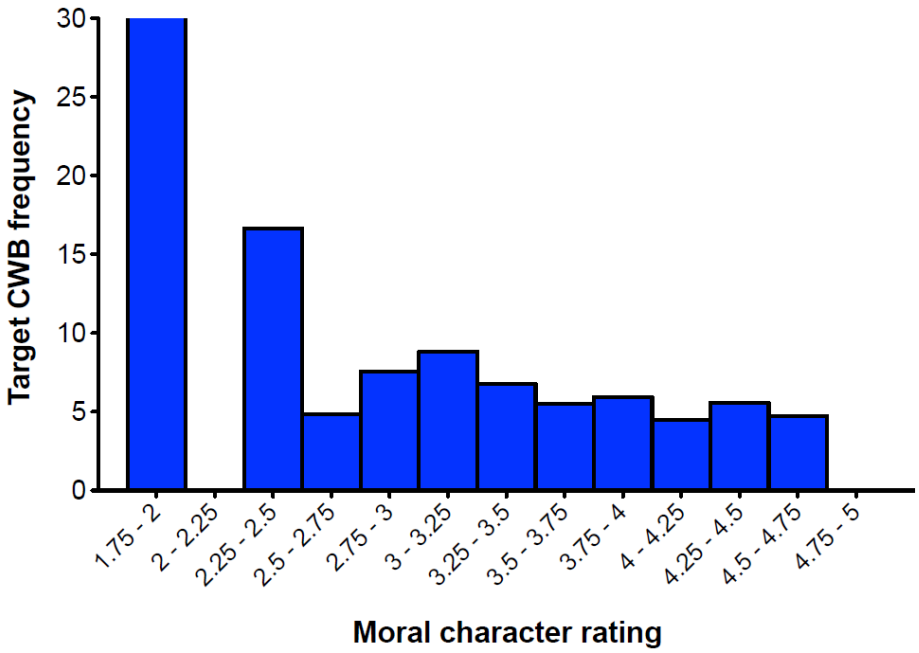

Figure 4. Study 2: Frequency of CWB as a function of Judges' Ratings of Moral Character for Dilemma Question

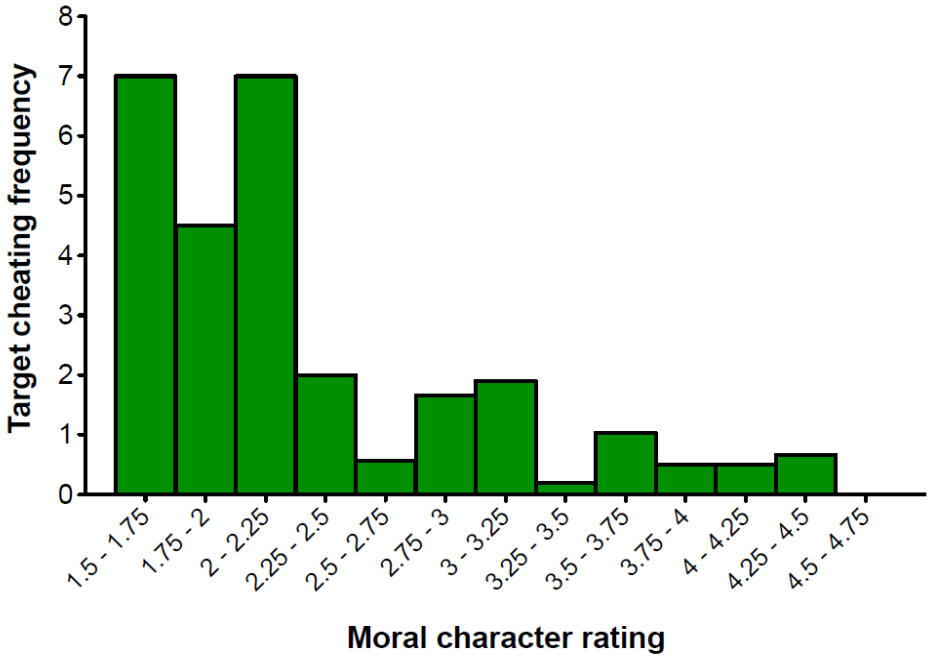

Figure 5. Study 3: Cheating in the Problem-Solving Task as a function of Judges' Ratings of Moral Character for Mistake Question

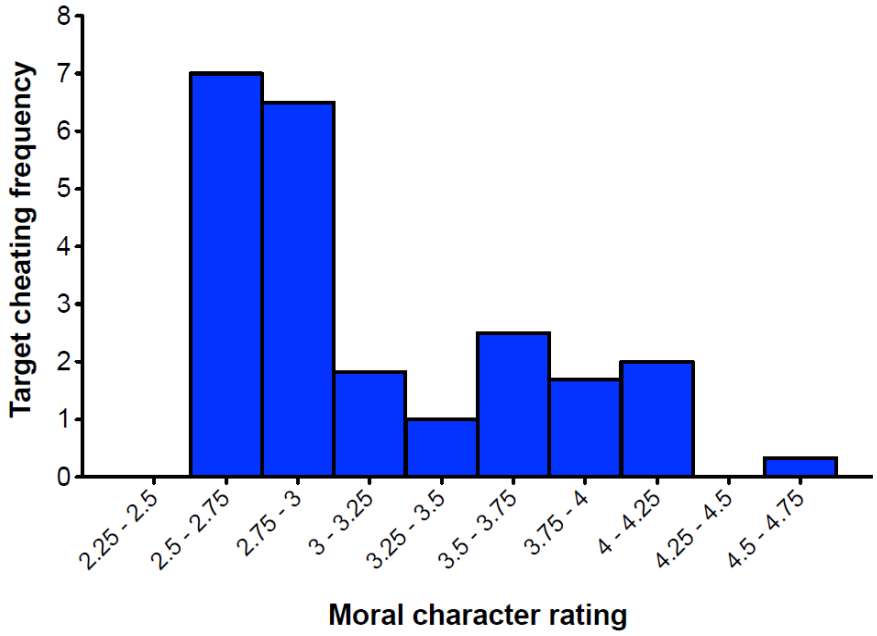

Figure 5. Study 3: Cheating in the Problem-Solving Task as a function of Judges' Ratings of Moral Character for Dilemma Question