

# **MORAL CHARACTER JUDGMENTS USING THE HIDDEN INFORMATION DISTRIBUTION AND EVALUATION (HIDE) MODEL**

YEONJEONG KIM

Tepper School of Business  
Carnegie Mellon University  
5000 Forbes Ave. Pittsburgh, PA 15213

TAYA R. COHEN

Carnegie Mellon University

A. T. PANTER

University of North Carolina at Chapel Hill

## **ABSTRACT**

We introduce the Hidden Information Distribution and Evaluation (HIDE) model, a new person perception framework. Applying the HIDE model to moral character judgments, we develop and validate a behavioral interview method to identify individuals who are likely to engage in unethical behaviors.

## **INTRODUCTION**

Moral character is an aspect of personality that describes an individual's disposition to think, feel, and behave in an ethical manner (Cohen, Panter, Turan, Morse, & Kim, 2014; Cohen & Morse, 2014; Fleeson et al., 2014; Kim & Cohen, 2015; Peterson & Seligman, 2004). 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. 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, Panter, Turan, Morse, & Kim, 2013).

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. This research is designed to remedy this dearth in our understanding of moral character judgments, especially when judges might need to make prompt evaluations from the limited information they obtain from strangers.

As a first step toward answering the question of how we can make accurate judgments of strangers' moral character, we develop the hidden information distribution and evaluation

(HIDE) model, which posits that evaluations made by targets and judges, respectively, capture different insights because some information is hidden from one party and is only detectable by the other. Applying the HIDE model to moral character judgments, we propose that judges who do not know the targets are able to detect aspects of moral character that target individuals incorrectly know 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. We propose that impromptu thinking and language usage in answering these questions reveal information about implicit moral character traits that targets are unable to hide but that judges can use to make accurate character judgments.

### **The Hidden Information Distribution and Evaluation (HIDE) Model**

At its highest level, the HIDE model separates person perception into two rating sources, self and judge. In the HIDE model, "judges" refers to people who might provide other-reports of targets, which could be targets' acquaintances or strangers who do not know the targets. The term "judge-report" is interchangeably used with other-report. The model assumes that there is information about the target's personality that can be judged (correctly or not) by the self, and there is information about the target's personality that can be judged (correctly or not) by others (i.e., judges). For each rating source, information about the personality of the target is distributed into three non-overlapping knowledge 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 (unknown or *hidden information*). An implication of viewing person perception through the lens of the HIDE model is that there are unique insights that one rating source (e.g., a judge) might have into a target's personality that the other party (e.g., the self) might not have insight into.

The correctly-identified-self component of the model (i.e., self-knowledge) is the knowledge that researchers aim to capture with self-reports. However, while self-reports are often an accurate reflection of the self and do remarkably well at predicting observable behaviors and important life outcomes (e.g., Ozer & Benet-Martinez, 2006; Roberts et al., 2007), they are not immune to errors and reporting biases, so they often capture invalid knowledge as well as valid knowledge. The invalid knowledge 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) (Paulhus, 1991). Self-deception refers to when targets have errors in their understanding of themselves. Impression management refers to when targets have accurate understanding of themselves but misrepresent that information, 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 knowledge. 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 at all in self-reports is the information in the hidden and hiding-self components of the model. The hidden-self component (self-ignorance) describes personality information that the target is unaware of and does not report on. The hiding-self component (self-screening) describes personality information that the target is aware of and decides not to report on. The combination of the incorrectly-identified-self, hidden-self, and hiding-self components of the HIDE model together 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 cannot provide.

Paralleling the self-report section of the model, the judge-report section of the model also shows that information about the target is distributed into three knowledge components: 1) valid information that the judge has about the target's personality (i.e., correctly-identified-target component; judge-knowledge), 2) invalid information that the judge has about the target's personality because of errors or bias (i.e., incorrectly-identified-target component), and 3) information about which judges do not report on because they do not know at all (i.e., hidden-target component; judge-ignorance) or they know of but decide not to report on (i.e., hiding-target component; judge-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 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 good. The combination of the correctly-identified-target and incorrectly-identified-target components capture the total 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.

### **Applying the HIDE model to Moral Character Judgment**

Trait evaluativeness is of chief importance when considering the relative value of self-reports versus judge-reports. Moral character is an inherently evaluative trait—people want to see themselves and be seen by others as highly moral. Distortion in self-reports increases as trait evaluativeness increases (Vazire, 2010; Vazire & Mehl, 2008). Using the language of the HIDE model, we can say that trait evaluativeness increases the amount of information that falls into the incorrectly-identified-self and hidden components as opposed to the correctly-identified-self component. That is, ego-protection and social desirability motives that stem from moral character being a highly evaluative trait are likely to increase self-deception and impression management, and thus push information about moral character into the incorrectly-identified-self zone of the HIDE model, and out of the correctly-identified-self zone. It follows, then, that judge-reports should be used to complement or replace self-reports to the extent that judge-reports can tap into valid information that is located in the incorrectly-identified-self, hidden-self, and hiding-self zones of the model.

Judge-reports of moral character can be provided by people who know the target well (i.e., well-acquainted others) or by strangers who do not have any social relationship with the target, but who have access to information relevant to judging targets' moral character. Often we assume that well-acquainted others will be better judges of personality than strangers, and studies generally support this claim. However, the HIDE model suggests that, in some circumstances, ratings made by strangers can be just as informative or even more informative than ratings made 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 of judges to correctly construe targets' moral character. Using the language of the HIDE model, judgments from strangers are likely to be more accurate than

judgments from well-acquainted others (e.g., friends) in circumstances in which well-acquainted others have incorrectly-identified-target knowledge (i.e., judge-error or judge-bias) but strangers do not.

## **Accuracy of Moral Character Judgments from Job Interview Questions**

The HIDE model posits that judges who do not know the targets are able to detect aspects of moral character that the target individuals misconstrue and/or are unaware of themselves. 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. An interesting and practical tool that judges might use in evaluating strangers' moral character is to ask open-ended questions that are designed to reveal information about targets' moral character. We propose that impromptu thinking and language usage captured in written responses to behavioral interview questions can reveal information about targets' moral character that judges can use to make accurate character judgments. To test the plausibility of this claim, we developed a battery of interview questions designed to covertly reveal people's moral character, and had judges who were unacquainted with the targets evaluate the targets' responses. We investigate the validity judges' moral character evaluations by measuring how well the evaluations predict the targets' unethical behaviors.

## **METHOD**

### **Interview Questions**

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 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*]
- 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*]

### **Data Collection from Targets**

In total, 195 U.S. adults participated in an experiment in a mobile research laboratory parked in the city of Pittsburgh, Pennsylvania. These participants answered one of the two interview questions, completed a problem-solving task in which they had the opportunity to lie about their performance, and answered a computerized survey in which they answered questions capturing demographic information as well as several personality traits.

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.

### **Data Collection from Judges**

One hundred and two participants were recruited from a university-administered subject pool to judge the targets' moral character (55.9% were female; the average age was 21.6, ranging from 18 to 69). 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**

In this study, the criterion variable, cheating, is operationalized as the number of matrices the participants claimed they solved minus the number they actually solved correctly. We formally tested the predictive validity of our interview method by conducting negative binomial

regression analyses. Targets' frequency of cheating was regressed on judges' moral character evaluation. The number of correctly solved matrices was controlled because participants who solved more matrices correctly had less opportunity to cheat. The regression results were similar, regardless of whether the Mistake and Dilemma questions were analyzed together or separately. Judges' moral character evaluation 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 ( $b = -.98$ ,  $S.E. = .25$ ,  $p < .001$ ) or Dilemma question ( $b = -.72$ ,  $S.E. = .32$ ,  $p < .001$ ).

## **DISCUSSION**

The HIDE model and its implications for moral character evaluations has the potential to make groundbreaking theoretical and applied contributions to organizational behavior 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. 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 reliably assess it.

Virtually all managers desire an ethical work force, yet little evidence-based guidance exists for assessing moral character. The current studies suggest that judges can make reasonably reliable and valid judgments of strangers' moral character based on short written responses to interview questions. Researchers could use the character-interview questions developed in this work 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.

Future research should examine more factors that play a role in the reliability and accuracy of moral character judgments in interview settings. Specifically, future research should investigate whether some judges are more accurate than others and whether some interview questions are more informative than others. It is also critical that future research examine how judges are trained and what they are asked to judge.

## **REFERENCES AVAILABLE FROM THE AUTHORS**