Moral Neutralization and Counterproductive Behavior

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Samuel D. Lee

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Nathan R. Kuncel, Advisor

2023

Acknowledgements

Thank you to my advisor, Professor Nathan Kuncel, for his feedback and guidance throughout these projects. I am also grateful for the expertise and thoughtful contributions of Paul Sackett, Deniz Ones, and Michelle Duffy. I am also grateful to Amy Riegelman, whose expertise and invaluable contributions to Study 2 undoubtedly saved me several hours of work. Thanks also to my fiancée Megan for her support and patience during the dissertation process.

Table of Contents

Introduction

Psychologists have long recognized the importance of rationalization. Psychological work on rationalization dates all the way back to Freud's early writings (Freud, 1901; 1914). Rationalization was viewed as "an operation that fulfills functions in the mental life independently of its degree of truth." (Zepf, 2011, p. 149). Even as many other Freudian ideas declined in popularity, Freud's broad understanding of rationalization persisted. Scholars continued to argue that rationalizations are used to avoid uncomfortable conflicts with reality (e.g., Bone, 1975; Sandler, 1976), and deceive the conscious self (Bibring et al., 1961; Freud, 1981; Levin, 1978). Research on rationalization seems to have made a resurgence in the late 1950s in response to psychology's cognitive revolution. For example, Festinger's (1957) groundbreaking theory of cognitive dissonance describes the ways in which rationalization processes can help to reduce feelings of dissonance when behaviors are inconsistent with held values and beliefs. Even the most prominent behaviorists were forced to acknowledge rationalization's importance (Skinner, 1965). Skinner argued people rationalize their behavior to avoid punishment. Around this time, interest in cognitive approaches to rationalization extended beyond psychology departments. In 1957, two criminologists, Gresham Sykes and David Matza, published their article, "Techniques of Neutralization: A Theory of Delinquency."

Neutralization Theory

The core idea of Sykes and Matza's neutralization theory (1957) is that people who normally subscribe to a moral view justify their unethical behavior via neutralization, a process during which the offender invokes rationalizing attitudes called "techniques of neutralization". The guilt associated with antisocial actions is "neutralized", and therefore the individual can maintain positive self-evaluations. Sykes and Matza argue that "It is by learning these techniques [of neutralization] that the juvenile becomes delinquent" (1957, p. 667). Sykes and Matza originally suggested that techniques of neutralization were implicated in the etiology of immoral behavior, but other scholars have pushed back on that idea by pointing out that it is difficult to rationalize behaviors before they occur (e.g., Maruana and Copes, 2005). However, a large literature suggests that some individuals are reliably more likely to invoke techniques of neutralization, and doing so is associated with a wide variety of deviant behaviors.

Although neutralization theory (sometimes referred to as "drift theory") was first published in the 50s, it has only grown in popularity and remains relevant today. Sykes and Matza's original article has been cited thousands of times and continues to accrue hundreds of citations each year. Additionally, although neutralization theory was originally proposed to explain the behavior of juvenile delinquents, it has been applied to virtually all domains. Scholars have collected empirical evidence linking neutralization to minor deviance, such as Sunday shopping among Mormons (Dunford & Kunz, 1973), moderate counterproductive behaviors, including stealing workplace property (Hollinger, 1991) or taking part in cyberbullying (Lowry et al., 2016), and even the most extreme immoral actions, such as becoming a hitman (Levi, 1981), rape (Bohner et al., 1998), and genocide (Alvarez, 1997; Bryant et al., 2017). Quantitative studies find strong associations between neutralization and counterproductivity. For example, Lee et al. (2020) found that neutralization was the single strongest individual difference correlate of academic dishonesty in higher education, with correlations exceeding those of traits like conscientiousness. Although neutralization theory is frequently applied to a wide range of settings, "it has found its most receptive audience in studies of organizational and white-collar crime" (Maruana & Copes, 2005, p. 223). Neutralization theory is frequently applied to the workplace (e.g., Kvalnes, 2019).

Sykes and Matza (1957) described five types of techniques of neutralization common among delinquents: denial of injury, denial of responsibility, denial of victim, appeal to higher loyalties, and condemnation of the condemners. *Denial of injury* is the tendency for the delinquent to justify a deviant behavior because they believe there is no significant injury that results from the behavior. *Denial of responsibility* allows the delinquent to mitigate guilt associated with performing a deviant act by believing that their actions were out of their control. *Denial of victim* is the tendency for the delinquent to believe deviant behaviors were acceptable because the victim deserved whatever happened to them. *Appeal to higher loyalties* occurs when the delinquent defends their deviant behaviors by appealing to the norms of a social group that they are a part of. Finally, *condemnation of the condemners* occurs when individuals who commit deviant behaviors focus on the hypocritical nature of the authority figures who are condemning the behaviors. In addition to the original five techniques, other techniques have been proposed. One of the most prominent and highly studied was proposed by Klockars (1974) who described the *metaphor of the ledger*, which is evoked when one feels they have a sufficient supply of good to their credit so can indulge in some evil without feeling guilty (Minor, 1981).

These six techniques of neutralization are those most commonly examined in the literature, but dozens more have been created. For example, Bryant et al. (2017) proposed appeals to good character and victimization. Appeals to good character occur when the individual who committed the deviant behavior asserts that they are incapable of committing antisocial behaviors or that they are acceptable because of their virtuous deeds and admirable qualities. Victimization occurs when the individual argues that the antisocial behavior is justifiable because they or people they are close to are under threat. Specific techniques tend to be more related to certain specific behaviors. For example, the reason Bryant et al. proposed appeals to good character and victimization is because they were especially relevant among perpetrators of genocide. However, due to relatively recent development and their context-specific application, limited research has been done on techniques beyond the original five and Klockar's (1974) metaphor of the ledger.

Attempts have been made to organize the many techniques of neutralization. Kaptein and Helvoort (2019) created a hierarchical taxonomy of neutralization techniques. They organized techniques into four broad categories: distorting the facts, negating the norm, blaming the circumstances, and hiding behind oneself. Each of the categories contains three sub-categories which each consist of five specific techniques. However, despite the existence of frameworks like Kaptein and Helvoort's, a majority of studies that measure neutralization techniques in some quantifiable way do so by summing self-report scores on the original five (or the original five plus the metaphor of the ledger). This is done because respondents who invoke one technique of neutralization are much more likely to invoke others. That is, scores on specific neutralization techniques are highly correlated (e.g., Curasi, 2013), so it is common practice to create an overall neutralization propensity score. Due to this tendency, neutralization is often treated as a dispositional trait-like variable. Empirical evidence provides support for a neutralization trait. For example, Jacobson (2020) found that situational factors, including stress and time-sensitive obligations, were mostly uncorrelated with neutralization, suggesting that neutralization tendency is a consistent disposition rather than merely a situationally driven attitude. A distinction must be made between invoking a technique of neutralization in a given situation and the general tendency to do so. Both are often referred to as simply "neutralization". For the remainder of this dissertation, I use the term “neutralization” to refer to the trait-like tendency to invoke neutralization techniques to justify antisocial actions.

The neutralization literature would benefit from improved measurement. When estimating a neutralization score, most researchers create a new measure. A few historically popular measures exist, but they were designed for juvenile samples. For example, the most popular classic neutralization measure is Ball’s (1966). This measure presents four scenarios (a gang fight, a gang fight with weapons, shoplifting, and theft with a weapon). Ten excuses for the behaviors are presented for each scenario, and respondents rate the extent to which they agree with each excuse. All forty items are summed to form the neutralization acceptance total score. Ball’s measure has been used often, but it typically goes through major changes when it is adapted to fit a new setting and sample. For example, Haines et al. (1986) adapted Ball’s measure in their study of academic dishonesty. However, the resulting 11-item measure applied to college student samples is so different from the original Ball measure that it can more accurately be described as a different measure altogether. Even when used as intended, Ball’s measure suffers from several major flaws. For example, Shields and Whitehall (1994) argue that the measure is "too long (in that many young offenders have short attention spans), too verbally sophisticated (given their limited vocabularies), and beyond the reading skills of many” (p. 227).

In most cases, researchers create new self-report measures to fit the needs of a single study. These measures are usually short, measuring each of the original five techniques of neutralization with only one or two items to form a sum/mean score, and tailored to specific outcomes. For example, the Haines et al. (1986) measure was designed to predict academic dishonesty, so the items ask respondents to rate how much they agree with excuses used to justify cheating. This is consistent with recommendations by researchers arguing that correlations can be maximized by targeting a neutralization measure to a specific outcome of interest (Maruana & Copes, 2005). However, an unfortunate consequence is that resulting measures can only be applied in narrow circumstances. Although such measures tend to produce reliable data when used appropriately, no measures have gone through extensive validation processes.

It is important to note that a substantial portion of moral neutralization research does not rely on self-report survey measures. The neutralization literature also includes an extensive qualitative literature that creates fertile ground for future hypothesis generation and theory development (e.g., Benson, 1985; Klenowski et al., 2011, Willot et al., 2011). This research examines court testimony, interviews, and writings to evaluate the presence and nature of neutralizing behaviors in a wide range of samples. The focus is on a general propensity to deny, neutralize, and redirect responsibility for immoral behavior. For example, in a highly cited study, Benson (1985) interviewed a variety of white-collar criminals, including antitrust violators, embezzlers, and other types of fraudsters. The criminals justified their crimes by stating that they were just doing what was necessary for the organization to remain competitive (the appeal to higher loyalties technique), by stating that the crime was benign, especially compared to more common criminal behavior (the denial of injury technique), and by invoking other rationalization strategies. Regardless of how neutralization is operationalized across studies, one thing remains clear. The tendency to invoke neutralization techniques to justify antisocial behavior is associated with counterproductivity across a wide range of situations.

Moral Disengagement

Decades after Sykes and Matza first published their neutralization theory, Albert Bandura began writing about a similar theory. Bandura (1990) originally proposed his cognitive theory of moral disengagement in an attempt to explain how terrorists and other extreme antisocial actors can live with the guilt associated with their abhorrent actions. Interestingly, this paper made no reference to the existing neutralization literature. which was active and 40+ years old while using the language of neutralization "Self-exonerations are needed to *neutralize* self-sanctions and to preserve self-esteem." (Bandura, 1990, p. 43, emphasis added).

Since that time, Bandura and other researchers have continued to develop his theory and apply it to counterproductive behavior of all types and intensities. According to moral disengagement theory, antisocial actors nullify the guilt that would be caused by their immoral actions via moral disengagement, which is the result of "the cognitive reconstruction of inhumane conduct into a benign or worthy [conduct.]" (Bandura, 1999, p. 193). Individuals do this by invoking rationalizing attitudes called "mechanisms of moral disengagement." Bandura (1996) described eight different types of mechanisms of moral disengagement that antisocial actors can invoke to justify their behaviors: moral justification, euphemistic language, advantageous comparison, displacement of responsibility, diffusion of responsibility, distorting consequences, attribution of blame, and dehumanization.

Moral justification occurs when detrimental conduct is made personally and socially acceptable by portraying it in the service of valued social or moral purposes. For example, a student athlete might justify cheating to maintain a GPA that allows them to remain on a team. Cheating would be deemed acceptable because it was for the good of the team. *Euphemistic language* occurs when sanitized and convoluted language is used to make destructive behavior appear benign. *Advantageous comparison* describes situations in which the antisocial actor compares their actions to more extreme immoral behaviors, thus rendering their actions benign in comparison. *Displacement of responsibility* describes a situation where people view their actions as a result of external forces that they are not responsible for, such as social pressure. People invoking displacement of responsibility often make statements like “I’m just following orders.” *Diffusion of responsibility* occurs when personal agency is obscured by spreading responsibility throughout a group. Responsibility can be diffused by division of labor for a venture with different members performing subdivided aspects that seem harmless in themselves but are harmful in their totality. For example, group decision-making could enable otherwise considerate people to behave inhumanely. When everyone is responsible, no one individual feels responsible. *Distorting the consequences*: when antisocial actors minimize or deny the harm caused by their actions. *Attribution of blame* occurs when the offender views themselves as a faultless victim who was forced to behave in a deviant manner by an outside, adversarial force. Finally, *dehumanization* occurs when victims are viewed as sub-human or have bestial qualities bestowed upon them. A list of all eight mechanisms of moral disengagement can be seen in Table 1.

Unlike in neutralization theory, scholars have largely focused on Bandura’s eight mechanisms and have not proposed many more. Indeed, it is relatively rare for researchers to focus on individual mechanisms. This may be because Bandura et al. (1996) recommend treating moral disengagement as a single broad construct. Additionally, multiple studies have found that a one-factor solution to moral disengagement measurement tends to fit best, even when all eight mechanisms are measured (Bandura et al., 1996; Detert et al., 2018; Moore et al., 2012). When mechanism-level subscores are provided, they are usually estimated using only three or fewer items each.

Like with neutralization, the general tendency to invoke such mechanisms of moral neutralization across a variety of situations can be conceptualized as a trait-like variable called “moral disengagement” or “the propensity to morally disengage”. Multiple popular self-report measures have been developed to measure moral disengagement in adults (e.g., Detert et al., 2008; Moore et al., 2012). They include items measuring all eight mechanisms of moral disengagement which are summed to generate a total score. The measures can be applied to adult samples across a variety of settings. Perhaps due in part to these high-quality and domain-general measures, research on moral disengagement in the workplace has exploded in popularity during the previous decade (see Newman et al., 2019 for a review). A large and growing body of research suggests that moral disengagement is a significant correlate of important workplace outcomes, including counterproductive work behaviors (CWB; Christian & Ellis, 2014), organizational citizenship behaviors (OCB; Bonner et al., 2016), turnover intentions (Nguyen, 2015), unethical pro-organizational behavior (Chen et al., 2016), job performance (Keem et al., 2018), and job satisfaction (Claybourn, 2010).

A Synthesis

Neutralization theory and moral disengagement theory are conceptually very similar. The theories converge on a nearly identical overarching concept: people rationalize and restructure how they think about their immoral behaviors to justify those behaviors and reduce their perceived responsibility and guilt. In some cases, this restructuring can even transform unethical behavior into a perceived virtuous action. Each theory described a non-exhaustive set of rationalization strategies individuals can employ. Several of these strategies are effectively identical across theories. For example, the neutralization technique “denial of injury” and the mechanism of moral disengagement “distorting consequences” both involve minimizing the perceived harm of an action. Even when a direct analog to a mechanism of moral disengagement is not present in the original five techniques of neutralization, a similar technique can usually be found in the literature. For example, the denial of humanity technique (Alvarez, 1997) is similar to the dehumanization mechanism. The techniques and mechanisms in Table 1 are ordered such that similar techniques and mechanisms are adjacent to each other when possible.

Empirical work supports the conceptual overlap of moral disengagement and neutralization. In a youth sample, scales of neutralization, moral disengagement, and self-serving cognitive distortion (another similar theory) intercorrelated between .51 and .64 (Ribeaud & Eisner, 2010). Ribeaud and Eisner also found that measures of both constructs loaded primarily on a single factor. Ribeaud and Eisner (2010) called this higher-order unified concept “moral neutralization.” Given that Neutralization appeared decades before moral disengagement, this *moral neutralizatio*n term will be used for the remainder of the dissertation.

Both theories are also similar in how they can be applied. For example, as mentioned previously, the neutralization literature is rich with qualitative interview studies. Benson (1985) interviewed several white-collar criminals and found that they invoked rationalizing attitudes to justify their crimes. The author interpreted these rationalizations through the lens of neutralization theory, but moral disengagement theory would have worked just as well. For example, criminals justified their crimes by stating that they were just doing what was necessary for the organization to remain competitive. Benson (1985) framed this as the appeal to higher loyalties technique, but the moral justification mechanism of moral disengagement applies in the same way. Similarly, criminals justified their actions by stating that the crime was benign, especially compared to more common criminal behavior. This could be interpreted as the denial of injury technique or the distorting consequences mechanism. Neutralization and moral disengagement are highly similar and can be applied in similar situations.

This similarity is made even more apparent when measures of neutralization and moral disengagement are compared at the item level. For example, the moral disengagement item from Moore et al. (2012) "It is okay to tell small lies when negotiating because no one gets hurt" is very similar in structure and wording to the Thurman (1984) neutralization item "It is okay to break the law if no one gets hurt" and the Siponen and Vance (2010) neutralization item "It is OK to violate company security policy if no one gets hurt." Not only is the method of rationalization illustrated in the items equivalent, the wording in the later portions of the items is identical. Many other item pairs are similar in content and structure across measures, such as Harris and He's (2019) moral disengagement item "It's alright to take things if you really need them" and Agnew and Peters' (1986) neutralization item "Would you feel guilty for shoplifting if you needed the item very badly but did not have the money to pay for it?"

Despite the similarity of neutralization and moral disengagement, both research streams have progressed largely independently. Even in Bandura's original work on moral disengagement, no mention was made of neutralization theory. When considering rationalization to justify antisocial actions, sociologists and criminologists cite neutralization theory whereas psychologists cite moral disengagement theory. As a result, much of the independent work done in both literatures is redundant. In the words of Howard and Levinson (1985, p. 191), the result is a "wasteful duplication of effort that follows from mutual interdisciplinary ignorance."

Only recently have researchers begun to aggregate outputs from moral disengagement and neutralization research. For example, as mentioned previously, Ribaeud and Eisner (2010) found that measures of moral disengagement and neutralization both load on a common moral neutralization factor. However, that research is limited in that both measures were meant to be used on adolescent samples and were developed specifically to predict aggression. The purpose of the present dissertation is to begin to understand the nature and utility of the moral neutralization construct when applied to adult samples.

Study 1: Structure of Moral Neutralization

Neutralization and moral disengagement measures each appear to tap into an underlying general moral neutralization construct (Ribeaud & Eisner, 2010). However, the little empirical work that has been done in this area has focused on moral neutralization as it relates to aggressive behaviors in children. Therefore, the extent to which the neutralization and moral disengagement items that are used in adult samples load on a general moral disengagement factor is unclear.

Researchers recommend that moral neutralization measures be tailored to specific outcomes (e.g., Maruana & Copes, 2005). As a result, measures are typically highly contextualized, meaning that they contain items that pertain to specific environments and types of moral decisions. They are designed to predict specific counterproductive behaviors, such as academic dishonesty (Haines et al., 1983; Shu et al, 2011), aggressive driving (Swan et al., 2017), shoplifting (Agnew & Peters, 1986), infidelity (Lisman & Holman, in press), or cyberloafing (Cheng et al., 2014). These measures ask respondents to share their tendency to invoke rationalizing attitudes to justify specific types of counterproductive behaviors. To my knowledge, no adult-oriented and outcome-agnostic neutralization measures have been developed, although the same cannot be said for moral disengagement measurement, where such measures have been developed (e.g., Detert et al., 2008; Moore et al., 2012). However, even general moral disengagement measures do not purport to assess techniques of neutralization. There is no sufficiently broad measure of moral neutralization that can be used on adult samples.

Due to the disparate nature of neutralization and moral disengagement assessments, it is necessary to evaluate several measures to determine the factor structure of moral neutralization measurement. This can be accomplished by collecting data from multiple instruments and comparing multiple confirmatory factor models. Several models could describe the relationships between neutralization and moral disengagement measures. One option is a single-factor model (Model 1; Figure 1). Although the mechanisms of moral disengagement and techniques of neutralization can each be viewed as separate constructs, they are positively correlated, and Bandura et al. (1996) argue that moral disengagement should be measured as a single broad construct. When developing their assessments, Detert et al. (2008) and Moore et al. (2012) both found that a single-factor model fit the data well, providing support for Bandura et al.'s claim. Additionally, measures of neutralization and moral disengagement designed to predict the same outcomes appear to tap into the same underlying construct (Ribeaud & Eisner, 2010).

If most neutralization and moral disengagement items are associated with the general tendency to invoke rationalizing attitudes to justify behavior, a single-factor model may fit well. However, the evidence in support of a one-factor solution comes from data collected from single measures (Bandura et al., 1996; Detert et al., 2008; Moore et al; 2012) or from data collected from multiple measures designed to predict the same type of outcome (i.e., childhood aggression; Ribeaud & Eisner, 2010). When data comes from many measures designed to predict different outcomes, results could be different. If the contextualized nature of moral neutralization measures accounts for a substantial portion of their variance, it is unlikely that a single-factor model will fit as well as an alternative model that accounts for the unique variance associated with each assessment. A similar potential model is a two-correlated factor model (Model 2; Figure 2), with one factor associated with moral disengagement items and another with neutralization items. This model is theoretically similar to the single-factor model and has many of the same strengths and weaknesses, but it would be more appropriate if neutralization and moral disengagement are indeed distinct constructs.

Another possible model includes X uncorrelated factors (Model 3; Figure 3), where X is the number of different moral neutralization measures included in the analysis (five in this case). Each factor would have an indicator for each item included in the associated measure. Such a model would likely fit well if the observed variance in moral neutralization measures could be attributed to their contextualized nature rather than due to broad underlying moral neutralization constructs. However, moral disengagement and neutralization theories suggest that the general tendency to invoke rationalizing attitudes should be correlated with counterproductivity even when measures are decontextualized. This is supported by research that finds that relatively decontextualized measures are still associated with counterproductivity (e.g., Bandura et al., 1996; Moore et al., 2012). Additionally, the findings of one study suggest that contextualized measures predict counterproductivity in different contexts. For example, academic dishonesty neutralization measures were correlated with CWBs (Lee & Kuncel, 2020). Therefore, it is unlikely that an X uncorrelated factors model would fit the data well.

A promising model is a one-factor bifactor model with a specific factor for each assessment included in the analysis (Model 4; Figure 4). A bifactor model is a multidimensional model that estimates at least one general factor and several uncorrelated specific factors. When estimating the bifactor model, each item included in the analysis exclusively loads on the general factor and one specific factor. In the present study, each item will load on the general moral neutralization factor and a specific factor associated with the relevant measurement instrument. The bifactor model provides multiple advantages. First, it allows the general moral neutralization factor to be estimated without it being entirely contaminated by the context-specific variance associated with the individual measures. This estimated general factor can be estimated and used to compute relationships between moral neutralization and other variables of interest. Similarly, the amount of variance in item responses accounted for by a specific factor rather than the general factor can be estimated, and the specific factors' correlations with outcomes can be estimated. This can provide a test of the impact of contextualizing the measures in terms of the respective behavioral outcome. For example, the neutralization measure developed by Haines et al (1983) was developed with the purpose of predicting academic dishonesty, and the wording of its items reflects that goal (e.g., "Jack should not be blamed for cheating if the instructor doesn't seem to care"). Therefore, I expect stronger specific factor loadings for Haines et al's measure, and if contextualizing the measure increases prediction, the specific factor should be correlated with academic dishonesty despite being uncorrelated with the general moral disengagement factor.

One final option is a two-correlated factor bifactor model (Model 5; Figure 5). This model has many of the same strengths as the one-factor bifactor model, but would be more appropriate if moral disengagement and neutralization are separate constructs. Each of the proposed models can be seen in Figures 1-5.

Neutralization and moral disengagement measures are theoretically similar, and previous research has found that they load on a common factor (Ribeaud & Eisner, 2010). However, adult assessments tend to contain items relevant to a specific criteria of interest. That is, the context of the measures can differ substantially. As a result, a substantial portion of the variance in scores on each of the measures can likely be attributed to that context rather than only the general moral neutralization construct. Therefore, I hypothesize that the one-factor bifactor model will fit data obtained from several measures better than the other four proposed models.

Hypothesis 1: Model 4 will fit the data better than the other four models.

A strength of bifactor models is that they allow us to estimate the general and specific factors separately. If either of the bifactor models fit the data adequately, the general factor(s) will be extracted and correlational analyses will be conducted to assess the strength of the association between moral neutralization and counterproductivity after accounting for the variance associated with the specific factors. Also, to evaluate the impact of matching moral neutralization item content to an outcome of interest, the specific factors will be correlated with academic dishonesty and CWB. Doing so will estimate the relationships after accounting for the common variance associated with the general moral neutralization factor(s). To maximize prediction, researchers recommend matching the item content of moral neutralization measures to the outcome of interest (Maruana & Copes, 2005). It follows that specific factors associated with highly contextualized measures will be more strongly correlated with matched outcomes. Scores on a specific factor associated with a contextualized measure like Haines et al.'s (1986) measure of neutralization as it relates to academic dishonesty should correlate more strongly with academic dishonesty than with other outcomes. Even when outcomes are not matched perfectly, I expect correlations to be stronger when the content and the outcome are conceptually similar. For example, Agnew and Peters (1986) developed a measure of neutralization designed to predict shoplifting behavior. Shoplifting is more similar to CWBs targeting the organization (CWB-O) than CWBs targeting people (CWB-I). Employee theft is a CWB-O behavior (Bennett & Robinson, 2000), so the correlation with CWB-O should be greater.

Hypothesis 2: The general moral neutralization factor(s) will positively correlate with counterproductive workplace behaviors, academic dishonesty, and infidelity intentions.

Hypothesis 3: The specific factors will correlate with counterproductive workplace behaviors and academic dishonesty such that correlations will be strongest when the content of the measure associated with the specific factor is matched to the outcome.

Hypothesis 3a: The Haines et al. (1986) specific factor will be more strongly correlated with academic dishonesty than with counterproductive workplace behaviors.

Hypothesis 3b: The Agnew and Peters (1986) specific factor will be more strongly correlated with CWB-O than CWB-I and academic dishonesty.

In addition to computing correlations with the general and specific factor scores, it is important to estimate the proportion of variance in the observed variables that can be explained by the different latent factors.

RQ1: What proportion of the variance in item scores is explained by the general factor(s) and the specific factors?

This question can be addressed by computing McDonald's (1999) coefficient omega (ω). ω is an estimate of the percent of variance explained by the latent factors in a model. When working with a bifactor model, ω will be equal to the estimate of variance explained by the general factors (ωH) plus the variance explained by all of the specific factors. These statistics can be used to estimate the proportion of variance accounted for that can be attributed to the common factors ($\omega H/\omega$) and the specific factors ($\omega - \omega H$) (Rodrigues et al., 2016).

Methods

Participants and Procedure

The total sample consists of 1,443 participants. Roughly half were recruited via Prolific, an online platform that recruits paid participants for online research (N = 738). 375 participants were recruited via Amazon Mechanical Turk (MTurk). The remaining participants were undergraduate students participating in research for extra course credit via the University of Minnesota psychology department's Research Experience Program (REP; N = 328). Participants were surveyed as part of three larger data collection efforts which took place in March 2020 (N = 360), December 2021 (345), and October 2022 (738; all Prolific). In all cases, the entire survey took respondents approximately 45-75 minutes to complete. All surveys were completed online. Recent research suggests that measures retain their validity even near the end of long surveys (Bowling et al., 2022). However, to ensure that fatigue effects were controlled for, the order of assessments included in the survey was shuffled for each participant. Additionally, attention check items were included throughout the survey. For example, "This is an attention check item. Please select 'Disagree.'" Respondents who failed at least one attention check item were removed from analyses but still received full compensation for their participation. REP respondents received two REP points (the standard rate for one hour). REP points can be exchanged for extra credit in undergraduate psychology courses. MTurk respondents received $7.25 for completing the survey. The estimated time to complete the survey was 60 minutes. After receiving feedback from the MTurk respondents, it was determined that the pay was not sufficient. Additionally, the Prolific survey was slightly longer (75 minutes). Prolific respondents were paid $14.86 ($11.88 per hour).

There are no firm rules regarding the necessary sample size to conduct a confirmatory factor analysis, though several rules of thumb have been proposed. Some are very lenient; Gorsuch (1983) recommended a minimum of 100 respondents. Comrey and Lee (1992) state that a sample of 100 is poor, but 200 is fair, 300 is good, 500 is very good, and 1,000 is excellent. Others make different recommendations depending on the number of variables included in the model (Kline, 1994). Everitt (1975) recommends at least 10 responses for each variable, whereas Hair et al. (1995) recommend a ratio of 20 to 1. Ultimately, the recommended rules of thumb are arbitrary and the one thing that most researchers can agree on is that a larger sample is better. The total sample size of 1443 meets most of the aforementioned thresholds of acceptability. Additionally, each factor will be estimated from several indicators (minimum of 5), which reduces the necessary sample size (Mundfrom et al., 2005).

Materials

Three neutralization and two moral disengagement measures were included in both data collections. All five full measures can be viewed in Appendices 1 - 5. The neutralization measures each assess a range of techniques of neutralization, including the original five, and the moral disengagement scales measure all eight mechanisms of moral disengagement.

Neutralization - Academic Dishonesty. Respondents completed Haines et al.'s (1986) 11-item neutralization measure focused on academic dishonesty (α = .965). All items are scored on a 1 - 5 Likert scale ranging from "Strongly disagree" to "Strongly agree." The measure includes items such as "Jack should not be blamed for cheating if… the instructor doesn't seem to care." and "Jack should not be blamed for cheating if… the people around him made no attempt to cover their papers." Internal consistency reliability was estimated using the total sample (α = .965).

Neutralization - Shoplifting. Agnew and Peters (1986) developed a short 5-item measure of neutralization focused on shoplifting. Each item was scored on a 1-3 Likert scale ranging from "Very guilty" to "Not guilty" (e.g., "Would you feel guilty for shoplifting if: The store owner was wealthy and wouldn't miss the item?"). Internal consistency reliability was estimated using the total sample (α = .881).

Neutralization - Violating the Law. The final measure of neutralization was Thurman's (1984) 7-item neutralization assessment. Items were scored on a 1-4 item Likert scale ranging from "Strongly disagree" to "Strongly agree." Thurman's measure is focused on violating the law, with items such as "It is not wrong to violate the law when the victim involved is a dishonest person." and "It is okay to break the law if no one gets hurt." Internal consistency reliability was estimated using the total sample (α = .873).

Moral Disengagement - Driving. The Driving Moral Disengagement Scale (Swan et al., 2017) is a 13-item measure that uses a five-point scale ranging from "Strongly disagree" to "Strongly agree." The scale measures the extent to which respondents invoke mechanisms of moral disengagement to justify poor driving behaviors like speeding and aggressive driving. The measure includes items such as “Honking the horn loudly is just a way of letting off frustration.” and “Some drivers deserve to be treated like the idiots they are.” Internal consistency reliability was estimated using the total sample (α = .871).

Moral disengagement - General. The Propensity to Morally Disengage Scale (Moore et al., 2012) is an eight-item assessment of moral disengagement. It includes one item for each of the eight mechanisms of moral disengagement. Items are scored on a seven-point Likert scale ranging from “Strongly disagree” to “Strongly agree.” Unlike the other measures included in the assessment, it was not designed to predict one specific type of counterproductive behavior. An example item is “People can’t be blamed for doing things that are technically wrong when all their friends are doing it too.” Moore et al.’s measure has been used extensively in workplace settings (Newman et al., 2019). Internal consistency reliability was estimated using the total sample (α = .829).

Counterproductive Workplace Behaviors. Self-reported CWB was measured with the 19-item Workplace Deviance scale (Bennett & Robinson, 2000). Respondents share how frequently they engaged in specific counterproductive behaviors, such as “Made fun of someone at work.” Responses are scored on a 1 - 7 Likert scale ranging from “Never” to “Daily.” In addition to producing an overall CWB score, the measure contains two subscales - interpersonal deviance (7 items) and organizational deviance (12 items). Organizational deviance consists of behaviors that are “directly harmful to the organization,” and interpersonal deviance consists of behaviors that are “directly harmful to other individuals within the organization” (Bennett & Robinson, 2000, p. 1). Some may be hesitant to rely on self-report measures of counterproductive behaviors, but their utility is evident (Berry et al., 2012). Self- and other-rated

CWBs are correlated, and self-reports tend to outperform other reports. Additionally, a respondent can report CWBs that they conducted in secret - behaviors which would be missed if relying on reports from other sources. Internal consistency reliability was estimated using the total sample (α = .881, .840, and .829 for overall CWB, CWB-O, and CWB-I, respectively).

Academic Dishonesty. Self-reported academic dishonesty was measured via the 12-item scale developed by McCabe and Trevino (1993). Respondents are asked to report how often they engage in specific cheating behaviors. Items are scored from 1 (Never) to 4 (Many times). As long as responses are anonymous, students will typically admit to cheating on self-report measures. Whitley (1998) found that over 70% of college students admit to cheating at least once during their academic career. MTurk and Prolific respondents who were not students were instructed to “think of the last time you were in school.” before completing the measure. Internal consistency reliability was estimated using the total sample (α = .913).

Infidelity Intentions. Self-reported infidelity intentions were measured via Jones et al.’s (2011) Intentions Towards Infidelity Scale (ITIS). Responses are scored using a Likert scale ranging from 1 (Not at all likely) to 7 (Extremely likely). The ITIS includes 7 items and typically takes less than a minute to complete. The ITIS includes questions about behavioral intention and likelihood rather than actual past behavior. For example: “How likely do you think you are to be unfaithful to future partners?” This focus on intent and likelihood was seen as a strength for the present study because respondents who are not currently in a relationship, or who have never been in a romantic relationship, can still respond to each question. Previous research has found strong correlations between ITIS scores and self-reported frequency of infidelity behaviors (Jones, 2009). Additionally, recent research has found a very strong correlation between ITIS scores and a measure of infidelity moral disengagement (r = .78; Lisman & Holman, in press). The ITIS was only administered to the Prolific sample (α = .811).

Analyses

Traditionally, confirmatory factor analysis (CFA) models are compared via fit indices such as RMSEA or CFI. However, these statistics can only be used to compare nested models. When models are nested. the parameters in the larger model but not in the smaller, nested model are being tested. The models proposed previously are not nested. Therefore, an alternative method of comparing the models was necessary. Models were compared by estimating the Akaike information criterion (AIC). AIC evaluates how well models fit the data by taking into account the number of parameters (by penalizing complexity) and how likely the observed data is given the model. The absolute value of the AIC is in part affected by an arbitrary constant defined by the data. Therefore, it cannot be interpreted. Instead, only differences in AIC values produced by the same dataset can be interpreted, with smaller values indicating better relative fit. Formally, the AIC function is $2K - 2ln(L)$ where K is the number of parameters in the model and L is the maximized likelihood function. Therefore, AIC will be smallest when the model fits the data well and has few parameters. Some have argued that AIC cannot be used to compare non-nested models, but this is a misconception (Anderson & Burnham, 2006). AIC only requires that the models were estimated using the same data to make comparisons. These analyses were conducted in R using the psych (Revelle, 2021) and lavaan (Rosseel, 2012) packages.

After estimating the models, the factor score indeterminacy of the bifactor models' factors were estimated using Guttman's indeterminacy index (Guttman, 1955), which is the correlation between two sets of factor scores that are minimally correlated with each other.

Additionally, coefficient ω was computed to estimate the proportion of variance in scale scores explained by the bifactor model's general factor and specific factors. Correlations between factor scores and criterion scores were also estimated.

Data came from three sources that may differ in important ways (REP, Amazon Mechanical Turk, and Prolific). Therefore, model four (the hypothesized best-fitting model) was tested for measurement invariance. Specifically, metric (AKA weak) and scalar (AKA strong) invariance were tested. Configural invariance occurs when the model takes the same form for each group. That is, the pattern of relationships is equivalent. Metric invariance is supported when the factor loadings are the same for each group. Scalar invariance is supported when the loadings and the intercepts are equivalent across groups. Metric invariance was tested by constraining the factor loadings to be equal for each group and comparing the fit of the model to the configural invariance model. To test for scalar invariance, the factor loadings and intercepts were forced to be equal across groups, and the fit of the model was compared to the fit of the metric invariance model.

Results

Across the five models, AIC values ranged from 162334 (Model 1) to 151782 (Model 3). See Table 2 for all AIC values. Although Model 4 was not the best fitting model as hypothesized, fit was deemed adequate so analyses continued. The factor loadings for Model 4 can be viewed in Table 3.

Coefficient ω was computed to estimate the proportion of scale variance accounted for by Model 4's general factor. Overall, the model accounted for 97% of the variance in scale scores (ω = .972). A majority of this (94%) was due to the general factor (ωH = .943). Relatively little

(6%) was accounted for by the specific factors (ω*S* = .055). Guttman's Indeterminacy Indices were also estimated for Model 4 (Table 4). Guttman's index for the general factor was relatively strong (.824), indicating that factor scores are fairly stable. However, index values were weak for some of the specific factors.

Correlations between factor scores and criterion scores were computed. The general factor scores were substantially and significantly correlated with CWB, academic dishonesty, and infidelity intentions (r = .44, .49, and .34, respectively). The correlation between the Haines et al. (1986) specific factor (S_3) and academic dishonesty was .17. Additionally, the Agnew (1986) specific factor (S_5) was correlated .10 with CWB-O but .00 with CWB-I. The remaining factor score and criterion correlations can be viewed in Table 5.

Measurement invariance was not supported (see Table 6). The fit of the model changed significantly when the factor loadings (metric invariance) and intercepts (scalar invariance) were forced to be equivalent across groups (REP, Amazon Mechanical Turk, and Prolific). The presence of measurement noninvariance suggests that the construct may have a different structure depending on the group.

Descriptive statistics for all variables subset by source (Prolific, MTurk, and REP) were also computed and can be viewed in Appendix 6.

Discussion

Hypothesis 1 was not supported. The one-general factor bifactor model (Model 4) fit the data fairly well but was not the best-fitting model according to the AIC values. The two-general factor bifactor model (Model 5) and the five-uncorrelated factor model (Model 3) both fit the data better, with Model 3 fitting best. These results are surprising, seeing as there are strong theoretical and empirical reasons to believe each moral neutralization measure included in the analyses should load on at least one common factor (Bandura et al., 1996; Moore et al., 2012, Ribeaud & Eisner, 2010). The highly contextualized nature of most of the measures likely contributes to the relatively strong fit of Model 3, but even with that in mind, a model with at least one general factor was expected to fit better because the measures were constructed with a higher-order construct in mind (not merely the context), and scores on each measure are intercorrelated. In addition to the contextualized nature of the items, another factor that could contribute to the stronger fit of Model 3 is the differing response scales. All items included in the analyses were scored on a Likert scale, but the number of response options and the response type differed across measures. For example, Agnew's (1986) measure of neutralization was scored using a 1-3 Likert scale ranging from Not guilty to Very guilty. Respondents were asked how guilty they would feel if they engaged in certain theft-related scenarios. Haines et al.'s (1986) measure of neutralization asks respondents to state how much they agree that another person should be blamed for engaging in cheating behaviors in differing scenarios, and responses are score using a 1-5 Likert scale ranging from 1-5. The potentially differing response processes (would *I* feel guilty versus should *someone else* be blamed) in combination with the different scale types could contribute to the fit of measure-specific factors.

The different measures also emphasized different techniques of neutralization or mechanisms of moral disengagement, even when they were based on the same underlying theory. For example, three of the thirteen items on Swan et al.'s (2017) measure of moral disengagement are meant to assess the dehumanization mechanism. Only one of the eight items on Moore et al.'s measure of moral disengagement is meant to measure dehumanization. This differential emphasis on lower order facets of the construct could contribute to the stronger fit of Model 3.

Although Model 3 fit the data best, the bifactor models fit the data almost as well. Additionally, the rich literature discussed previously strongly suggests that different measures of moral neutralization are measuring at least one broad, general factor. Therefore, I will move on from Model 3 and continue by examining the bifactor models (Models 4 and 5).

The bifactor models fit the data similarly well, though the fit of Model 5 was slightly better. The difference between the two models may not be especially meaningful. A composite of the observed moral disengagement scores is strongly correlated with a composite of the observed neutralization scores ($r = .66$). This strong relationship is reflected in Model 5. The general factors in Model 5 were strongly correlated with each other ($r = .68$), suggesting that even if neutralization and moral disengagement measures are measuring different higher-order constructs, they are extremely similar. It may be that the small differences are the result of superficial differences, such as different item stems, response options, and superficially distinct mechanisms and techniques (rationalization strategies). Therefore, although it was not the best-fitting model as hypothesized, I felt comfortable using Model 4 in the analyses.

The estimated factor scores from Model 4 correlated with the criteria as hypothesized (supporting Hypotheses 2 and 3). The general factor, which is meant to be an estimate of moral neutralization after removing the specific variance associated with each measure, was substantially correlated with all three outcome variables (Table 5). This suggests that the correlations between moral neutralization and outcomes are not merely the result of their specific, contextualized nature. That being said, scholars argue that measures should be contextualized because prediction will be increased (Maruana & Copes, 2005). This is not unique to moral neutralization. Domain specific measurement of broad individual differences variables often results in increased correlations (e.g., Shaffer & Postlethwaite, 2012). Therefore, I hypothesized that the specific factors, which should include item context-specific variance, will be at least slightly correlated with matched outcomes. The results support this. Although most of the correlations between the specific factors and the outcome variables were near 0, specific factor 3, which draws variance from Haines et al.'s academic dishonesty neutralization measure, was correlated with academic dishonesty (r = .17). Additionally, specific factor 5 is based on items from Agnew's (1986) shoplifting neutralization scale. I argued that shoplifting is conceptually similar to CWB-O, so specific factor 5 should correlate slightly with CWB-O, but not other outcomes. That hypothesis was somewhat supported. Specific factor 5 was slightly correlated with CWB-O (r = .10), but only slightly stronger than with other outcome variables (see Table 5).

Altogether, the correlations between the predicted factor scores and the outcome variables were mostly as hypothesized. However, the correlations with the specific factors were quite weak. Additionally, the correlations between the specific factors and other variables must be interpreted with caution. Guttman's indeterminacy index was computed to estimate the stability of the factor scores. For each factor, Guttman's index is equal to the correlation between two minimally correlated sets of possible factor scores (Guttman, 1955). The value for Model 4's general factor was strong (.824), suggesting that factor scores and their relationships with other variables should be fairly similar across estimates. However, some of the specific factors resulted in lower values (Table 4), indicating a large degree of indeterminacy. Therefore, specific factor scores and the correlations with the specific factors, including those with academic dishonesty and CWBs discussed above, may be unstable and should be interpreted with caution.

To further investigate the factor scores, McDonald's (1999) coefficient ω was computed. 94% of the explained scale score variance was attributed to the general factor, and only 6% was explained by the specific factors. This suggests that the general factor is a larger determinant of scores of the measure than the specific factors, which theoretically reflect the contextualized item content. This may help to explain why correlations between the specific factors and outcomes were weak, even when context was matched. Contextualized items seem to produce stronger correlations, but the majority of item score variance is due to the broader moral neutralization factor.

The measurement noninvariance findings suggest that the structure of moral neutralization may be different depending on the subsample. However, this may not be a problem with moral neutralization, broadly. Instead, measure-specific factors and limitations of the study design may be introducing systematic error that is producing measurement noninvariance. For example, it is possible that REP respondents provided less thoughtful responses. REP participants were guaranteed to receive REP points for their participation regardless of the quality of their responses. However, on Amazon Mechanical Turk and Prolific, researchers can reject responses and prevent the respondent from being paid. Additionally, if they are consistently flagged as poor respondents, they are provided with fewer paid surveys. This difference in thoughtful responding is supported by a cursory review of the raw data. Only 1.2% of Prolific respondents failed at least one of the two attention check items. 4.5% of MTurk respondents failed at least one attention check item, and 10.5% of REP respondents failed at least one attention check item. These differences in thoughtful responding may contribute to the observed measurement noninvariance.

The contextualized nature of the moral neutralization measures could also contribute to measurement noninvariance. For example, the Haines et al. (1986) measure of academic dishonesty neutralization may be inappropriate for non-student samples. Respondents who were not students were asked to think back to when they were last in school. However, it is likely that most of the MTurk and Prolific respondents had not been students for many years. The average age of MTurk and Prolific respondents was 40.18 and 39.00 (Appendix 6), respectively, so they may no longer be as familiar with the academic context as the REP respondents, who had an average age of 20.47 and are necessarily enrolled in college. Additionally, with the rise of online cheating websites and artificial intelligence, the nature of academic dishonesty has changed substantially over the past decade. A similar issue may be affecting the scores on Swan's (2017) driving moral disengagement scale. Many students at the University of Minnesota do not have a car, so they may not be as familiar with driving norms and behaviors. A participant does not need to know how to drive to complete Swan's scale, but relevant knowledge could reasonably be expected to affect how the respondent thinks about each item. Using these different measures of moral neutralization in the same model could be the cause of the observed measurement noninvariance.

Overall, although the results are more nuanced than hypothesized, the findings provide support for a broad, higher-order moral neutralization construct that encompasses both moral disengagement and neutralization. Additionally, moral neutralization appears to be associated with important counterproductive outcomes. In the next study, a meta-analysis is conducted to learn more about the magnitude of these relationships and how they vary across settings and situations.

Study 2: Meta-Analysis

Both neutralization and moral disengagement are powerful predictors of dishonest behavior. In a previous meta-analysis, I found that neutralization was the single best individual difference predictor of academic dishonesty with correlations exceeding those of traits like Conscientiousness (Lee et al., 2020). However, although "[n]eutralization theory has found its most receptive audience in studies of organizational and white-collar crime" (Maruana & Copes, 2005, p. 223), work in organizational psychology has focused primarily on moral disengagement. In a systematic review, Newman et al. (2019) highlighted the links between moral disengagement and several undesirable outcomes, including counterproductive work behaviors (CWBs) and unethical pro-organizational behaviors. Neutralization is also frequently examined in the workplace, though not usually by psychologists. Quantitative studies typically examine specific counterproductive outcomes, such as security policy violations (Siponen & Vance, 2010), rather than a broad CWB construct.

Although there are many studies of neutralization in the workplace, several of them are qualitative interview studies and therefore cannot be included in meta-analyses. These qualitative studies find links between neutralization and counterproductive behaviors at work (e.g., Benson, 1985). Criminals who are interviewed invoke rationalizing attitudes to justify or downplay their deviant behaviors. However, the qualitative nature of the research means that correlations cannot be extracted and included in the present meta-analysis. The counterproductive behaviors found in these qualitative studies, such as embezzlement, also tend to be much more extreme than those found in quantitative studies which measure behaviors with a base rate sufficient enough to compute meaningful correlations. Far more moral disengagement studies use quantitative methods, and as a result, there will be a disproportionate amount of moral disengagement studies included in a meta-analysis of moral neutralization. However, it is clear from the literature and Study 1 that both moral neutralization constructs are strongly associated with counterproductivity.

Moral neutralization is associated with a variety of important workplace outcomes that may be of interest to organizational psychologists, including counterproductive workplace behaviors (e.g. Moore et al., 2012), job performance (e.g. Bonner et al., 2014), job satisfaction (e.g. Zheng et al., 2019); organizational citizenship behaviors (OCBs; e.g. He et al., 2019), turnover intentions (e.g. Christian & Ellis, 2014), unethical pro-organizational behaviors (UPBs; e.g., Umphress et al., 2020), and academic dishonesty (Lee et al., 2020). In their narrative review of the moral disengagement literature pertaining to the workplace, Newman et al. (2019) concluded that “moral disengagement exerts a significant influence on the work attitudes and behavior of employees at the individual level of analysis.” In a recent meta-analysis, Ogunforwora et al. (2021) found that moral disengagement had robust relationships with a variety of workplace outcomes. However, their analysis did not include studies on neutralization.

As discussed previously, moral disengagement and neutralization are conceptually extremely similar, a model with a general factor that includes measures of both fits well, and measures of moral disengagement and neutralization are strongly correlated with each other (r = .66; Study 1). Despite the evident importance of moral neutralization and the similarity of moral disengagement and neutralization, both theories have been researched largely independently. Additionally, observed relationships with outcomes are quite large, with correlations exceeding

.4 with counterproductivity occurring regularly across both literatures (e.g., KiYoung et al., 2016; Moore et al., 2019). Therefore, the goal of the present meta-analysis is to estimate the relationship between moral neutralization and workplace outcomes by synthesizing findings from both the neutralization and moral disengagement literatures. By doing so, I hope to bridge the gap between the moral neutralization theories and illustrate the importance of moral neutralization in the workplace.

Variables of Interest

Given that moral disengagement and neutralization were developed to explain counterproductive behavior, it is unsurprising that the most common individual-level workplace outcomes examined in the moral neutralization literature are CWBs. The relationships between moral neutralization and other important workplace outcomes have also been estimated in several studies (e.g., Christian & Ellis, 2014; Zheng et al., 2019). In the present meta-analysis, I estimate the relationship between moral neutralization and CWBs, OCBs, unethical pro-organizational behaviors, turnover intentions, and job satisfaction. I will also estimate the relationship between moral neutralization and academic dishonesty. Multiple studies suggest that each of the six aforementioned outcome variables is correlated with moral neutralization. However, there is variation in the magnitude of the relationship across studies. Therefore, meta-analytic methods will be useful for determining the average effect size and how much the effect varies across different settings and situations.

Academic Dishonesty

Academic dishonesty is not a workplace construct, but it is moderately correlated with CWBs (See correlations in Studies 3 and 4), and deviant actors in both academic and workplace contexts tend to go through similar decision-making processes (Harding et al., 2004; 2006). As a result, researchers of one topic frequently generate hypotheses based on findings and theories from the other (e.g. Giluk & Postlethwaite, 2014). Academic dishonesty is relevant to workplace outcomes and is a common variable in moral neutralization research, so it will be included in the meta-analysis.

In a previous meta-analysis conducted by Lee et al. (2020), neutralization emerged as one of the strongest individual difference correlates of academic dishonesty (k = 19, $\bar{\rho}$ = .43). The authors theorized that high levels of neutralization enabled students to overcome obstacles that may have otherwise prevented students from cheating. “[a]lthough a student might feel that cheating is wrong… and probably not necessary… neutralization sets up cheating behavior anyway.” (Lee et al., 2020, p. 1051). Lee et al. did not include moral disengagement in their analysis, but studies on moral disengagement have found similarly strong correlations with academic dishonesty (e.g., Fida et al., 2020). Additionally, due to the similarity of neutralization and moral disengagement, I anticipate similar results. A strength of including academic dishonesty in the analysis is that it has been a popular criterion variable in both moral disengagement and neutralization research. Therefore, the analyses will not be dominated by studies of only one type of moral neutralization.

CWBs

According to social cognitive theory (Bandura, 1986), individuals exercise cognitive control over their behaviors and thoughts via self-regulatory processes. A person’s self-regulatory processes inhibit antisocial tendencies and encourage the person to behave in a manner consistent with personal and societal moral standards. However, “there are many processes by which self-sanctions can be disengaged from inhumane conduct… Selective activation and disengagement of internal control permits different types of conduct with the same moral standards." (Bandura, 1991, p. 71-72). Individuals with a stronger tendency to invoke rationalizing attitudes will be more likely to cognitively dismiss self-sanctions, so they will be uninhibited and experience less guilt after behaving unethically. Therefore, I expect CWBs to be strongly associated with moral neutralization. Based on a review of correlational research on individual differences, Giganc and Szodorai (2016) concluded that correlations of .10, .20, and .30 are relatively small, medium, and large effects, respectively. Researchers have somewhat consistently found that the correlations between moral neutralization and counterproductivity exceed .30 (e.g. Lim, 2002; Meng et al., 2014). Therefore, I hypothesize that counterproductive workplace behaviors will be strongly correlated with moral neutralization.

Unethical Pro-Organizational Behaviors

Unethical pro-organizational behaviors can be defined as actions that are intended to promote the effective functioning of the organization or its members and yet also violate core societal values, morals, laws, or standards of proper conduct (Umphress & Bingham, 2011). An employee may be faced with a situation where they must choose between behaving ethically and behaving in a way that benefits their organization or their coworkers. Theoretically, this behavior is distinct from CWBs. Sackett and Devore (2001) defined CWBs as those behaviors that violate the legitimate interests of the organization. Law breaking and violating codes of conduct are not in the legitimate interest of an organization yet may be a strong aspect of the organization's culture. Conceptually, moral neutralization has a potentially interesting relationship with unethical pro-organizational behavior in that the rationalization for the behavior, helping the organization, is built into the construct. In accordance with the propositions of social cognitive theory (Bandura, 1986; 1990; 1991), individuals may rationalize their unethical behavior such that they can assist their organization or coworkers without guilt or self-disapproval. I expect to find a strong correlation between moral neutralization and unethical pro-organizational behavior.

OCBs

OCBs can be defined as the set of behaviors that sustain or enhance the cooperative system of the organization but are not systematically or generally recorded in the formal system of the organization or tied in any consistent way to specific rewards (Organ, 2018). The relationship between moral neutralization and OCBs is also theoretically interesting. Although they are correlated, OCBs are distinct from reverse-scored CWBs (Dalal, 2005). Failing to engage in extra-role behavior is not the same as behaving counterproductively, but rationalizing attitudes could still be employed to mitigate guilt associated with failing to help coworkers and the organization. Therefore, I expect that OCBs will be negatively correlated with moral neutralization.

Job Satisfaction

Social exchange theory posits that an employee's attitudes and behaviors depend on their relationships with their coworkers, supervisors, and organization as a whole (Blau, 1964). When the quality of those relationships is poor, an employee may be dissatisfied with their job. Poor organizational relationships can also provide additional opportunities for individuals to invoke rationalizing attitudes (Bandura, 1999). Therefore, I predict that moral neutralization will be negatively correlated with job satisfaction.

Turnover Intentions

I hypothesize that moral neutralization will be positively correlated with turnover intentions. Individuals with poor organizational relationships will be more likely to have turnover intentions and be more likely to invoke rationalizing attitudes (Huang et al., 2016).

Moderators

Meta-analytic methods can be used to measure the impact of moderating variables on the effect of interest. In cases where at least 20 studies are included in the meta-analysis (CWBs and academic dishonesty), moderator analyses will be conducted. A brief description of each moderator of interest is described below.

Measure

A key moderator of interest is the comparative predictive power of different operationalizations of the construct. Three moral disengagement measures, Bandura et al. (1996), Detert et al. (2008), and Moore et al. (2012) are the most common measures of moral disengagement used in workplace studies. There does not appear to be one popular self-report measure of neutralization in adults, so studies often create their own, or they adapt other measures to fit specific scenarios. Additionally, there have been far fewer quantitative studies of neutralization than moral disengagement on employee samples. Therefore, all neutralization measures will be grouped into one category for the moderator analysis. Each instrument to be included in the analysis should measure the same underlying moral neutralization construct, so I do not anticipate large moderating effects. However, each measure differs slightly based on how they are intended to be used. Neutralization measures are typically targeted towards a specific counterproductive behavior, which can result in stronger prediction (Maruana & Copes, 2005), whereas the most common moral disengagement measures are less contextualized. Additionally, a previous review found that the measure developed by Moore et al. (2012) performs especially well in employee samples (Newman et al., 2019). Therefore, I anticipate that the Moore et al and neutralization measures will yield the strongest correlations with CWBs.

Neutralization versus Moral Disengagement

I also examine the theory (neutralization versus moral disengagement) that the measure is based on as a moderator, but because of strong empirical, conceptual, and item overlap, I do not hypothesize moderating effects of theory.

Content Matching

Many measures of moral neutralization are contextualized such that their item content is matched to the criterion of interest. For example, Shu et al.'s (2011) measure of moral disengagement and Haines et al.'s (1986) measures of neutralization both assess rationalizations as they relate to academic dishonesty. Researchers have suggested that the measures perform best when tailored to the criterion of interest. As such, I hypothesize that correlations will be greatest when contextualized measures are used to predict corresponding criteria. The data used to test this hypothesis will overlap considerably with the data used to evaluate the effect of the measure. However, this analysis will allow me to include contextualized measures that are only used in one or two studies.

CWB Type

Given evidence that the type of CWB moderates the correlation with some individual differences (e.g., Berry et al., 2007) I also evaluate interpersonal and organizational CWBs as a moderator. Within the Five Factor Model, interpersonal traits like Agreeableness appear to be more strongly related to interpersonal CWBs. Moral neutralization tactics can be focused on other individuals or groups while others target society or organizations. As such, no clear construct matching is apparent with overall moral neutralization scores. Therefore, I do not expect CWB type to moderate the effect but explore it as a moderator due to its availability.

CWB Specificity

Finally, I also examine the specificity of the CWB outcome as a moderator. Many studies assess overall CWBs as the criterion of interest. For example, a common measure of the dependent variable is Bennet and Robinson's (2000) measure of workplace deviance, which assesses a wide range of CWBs. Other studies focus on more specific counterproductive outcomes, such as cyberloafing (Betts et al., 2014) and social undermining (Duffy et al., 2012). Moral neutralization is associated with a wide range of counterproductive behaviors, and I have no theoretical basis for suspecting a difference based on the specificity of the criterion, so I do not expect CWB specificity to moderate the correlation.

Other Moderators - Not Included

Other moderators were of interest but could not be included due to limited variability across studies. For example, previous research has found that the correlations with CWBs can vary depending on how CWBs are measured (e.g., Dilchert et al., 2007). However, nearly every study included in the analysis measured CWBs via self-report. None used objective measures. Another moderator of interest was whether or not participants in CWB studies were already employed or job applicants. If an employee engages in many counterproductive behaviors on the job, they would be more likely to be fired. Therefore, by including only current employees in a study, scores from the low end of the CWB score distribution may be missing. Due to this restriction of range, correlations computed in these studies may be weaker than those that would be observed in the full population. Unfortunately, all studies included in the analyses sampled only current employees, so this moderating effect could not be tested. Moderators were only tested when there were at least ten effects in at least two conditions.

Methods

Literature Search

A literature search of articles related to moral neutralization in the workplace was conducted to obtain data for this meta-analysis. Electronic searches were completed to locate studies of moral neutralization that included correlational analysis with at least one of the five workplace outcome variables described previously. Databases and search strings were selected in consultation with a university librarian. Broad database searches were conducted with Business Source Premier, Sociological Abstracts, Criminal Justice Database, and PsychInfo. Search terms included *Neutralization* and *Moral Disengagement* in combination with *Employee, Manager, Organization,* and *Workplace*. See Appendix 8 for complete search strings used to search each database. Unpublished studies were not specifically sought out, but they were included if they came up in the search and satisfied the inclusion criteria. After the database searches, the bibliography sections of key flagged articles were searched for additional studies.

The initial search was conducted in May of 2020. However, several studies were published shortly after the search, including a meta-analysis of moral disengagement correlates (Ogunfowora, 2021). In an attempt to collect more recent records, a qualitative literature search was conducted in the fall of 2022. Additionally, the references of Ogunfowora et al. were screened. As per the PRISMA guidelines (Moher, 2009), the number of records identified in each database and each other step of the search process was recorded and presented in Figure 6. Ultimately, 79 records were included in the workplace outcome analyses.

Another literature search was conducted during the fall of 2022 to update Lee et al.'s (2020) academic dishonesty meta-analysis with moral disengagement studies. Database searches were conducted with PsycINFO, Academic Search Premier, Education Source, Education Resources Information Center, OpenDissertations, JSTOR, and the top 60 results in Google Scholar after sorting by relevance. Databases were searched using a combination of the following keywords: *Moral Disengagement*, *Academic Dishonesty*, *Cheating*, and *Academic Cheating*. See Appendix 9 for complete search strings used to search each database. A total of 35 records were included in the academic dishonesty analyses. The number of records identified during each step of the search is presented in Figure 7.

Inclusion and Exclusion Criteria

To be included in the analysis, the study had to include correlations between neutralization or moral disengagement and one of the previously described outcome variables. If a study did not present a correlation but contained statistics that could be converted into a correlation (e.g. Cohen's d), then the study was included in the analysis. To increase power, I took a broad approach when including studies examining the correlation between moral neutralization and CWBs. If a study examined CWBs, unethical workplace behaviors, organizational or interpersonal deviance, or specific counterproductive workplace behaviors, the study was included and the outcome variable was treated as CWB for the sake of the analyses. I examine the differences between types of CWB via moderator analyses. To remain consistent with Lee et al. (2020), academic dishonesty studies were only included in the analysis if the sample contained college students. High school cheating studies were excluded from the meta-analysis. A link to the meta-analytic dataset can be found in Appendix 13.

Meta-Analytic Methods

Analyses were conducted in R using the Psychmeta and metafor packages (Dahlke & Wiernik, 2018; Viechtbauer, 2010). All key variables were treated as continuous in our analyses. Therefore, correlation was the unit of analysis rather than another metric of effect size (e.g. Cohen's d). For each variable, the most common forms of measurement are self-report measures with generally moderate reliability. As a result, I produced estimates of the sample-size weighted mean correlation corrected for measurement error in measures of both variables. To conduct the analyses, I used the Hunter and Schmidt (2004) random effects meta-analytic technique. Research comparing the Hunter and Schmidt (2004) meta-analytic method with other methods has found that they yield similar results (Schulze, 2004). The Hunter and Schmidt method includes methods for adjusting correlations for unreliability. This was done via the correction for attenuation formula:

$$r_{x'y'} = \frac{r_{xy}}{\sqrt{r_{yy} r_{yy}}}$$

where $r_{x'y'}$ is the corrected correlation, r_{xy} is the observed correlation, and r_{xx} and r_{yy} are the reliabilities of the measures. If a study did not report the reliability of a measure, the average reliability of measures of that variable was imputed. This deviates from the method used by Lee et al. (2020). Lee et al. used the Spearman-Brown prophecy formula (Brown, 1910; Spearman, 1910) to estimate the mean reliability of a measure after adjusting for its length. However, the Spearman-Brown formula assumes that all items are of equal quality, which may not be the case. Some of the estimated reliability coefficients in Lee et al. (2020) were quite small. Therefore, to avoid potentially overcorrecting, I imputed the overall average reliability of the construct's measures. Although not reported, analyses were conducted again using alternative methods of imputing reliability (Spearman-Brown, assuming perfect reliability), but because most studies reported reliability, different imputation methods produced negligible differences in the mean reliability-corrected correlation.

In some cases, correlations with multiple components of a construct of interest were reported, but not the correlation with the overall scores. For example, Sahi and Ahmad (2019) reported moral disengagement correlations with CWB-O and CWB-I, but not CWB generally. In cases such as these, it is common to average the correlations with the components, but this typically underestimates the overall correlation with the composite. Instead, I used the formula for computing the correlation between a composite and an outside variable from Ghiselli et al. (1981):

$$r_{ZC} = \frac{r_{oi}}{\sqrt{1/k + (\frac{k-1}{k})r_{ii}}}$$

Where r_{ZC} is the correlation between the composite and the outside variable, r_{oi} is the average correlation between the outside variable and each component variable, r_{ii} is the average correlation between the components of the composite, and k is the number of components. Similarly, to estimate the reliability of the composite, I used Nunnally and Bernstein's (1994) formula 7-15 for estimating the reliability of a composite:

$$r_{yy} = 1 - \frac{k - \Sigma r_{ii}}{\boldsymbol{\sigma}_y^{2}}$$

Where r_{yy} is the reliability of the composite, k in the number of components in the composite, $\sum r_{ii}$ = the sum of the component reliabilities, and $\boldsymbol{\sigma}_y^{2}$ is the variance of the linear composite, which can be estimated by summing the values in the correlation matrix.

Tables 7 - 10 include the sample size weighted mean observed correlation $\bar{r}$ and its standard deviation SD_r. Additionally, the sample size weighted, measurement error adjusted mean correlations where each correlation was adjusted for unreliability in the measurement of both variables $\bar{\rho}$ is presented. SD_{ρ}, the standard deviation of correlations after taking into account sampling error and the influence of differences in measurement error is also presented. To interpret the variability of the effects, 80% credibility intervals were computed by using SD_{ρ} around $\bar{\rho}$. Credibility intervals include the range of values that the correlation might take across different situations and settings. Wider credibility intervals are indicative of moderators and suggest that a given study's findings may not be generalizable (Schmidt & Hunter, 1977; Whitener, 1990). As another estimate of heterogeneity, I^2 was also computed. I^2 Is an estimate of the percentage of heterogeneity in effects across studies that is not due to sampling error.

Publication Bias Methods

One threat to the validity of meta-analytic findings is publication bias. Publication bias occurs when effects that a reviewer is able to collect differ systematically from the full universe of effects (McDaniel, 2009). A common explanation for why publication bias could occur is that statistically insignificant findings could be suppressed. Because journals are less likely to publish null results, the effect sizes found in the published literature are inflated, with the smallest published studies producing the strongest effect. However, this may not be a major problem for the moral neutralization literature examined. First, the majority of moral neutralization studies reviewed are not underpowered. The vast majority had a power of 80% or greater to find a correlation of .30, a very conservative estimate of the moral neutralization and counterproductivity relationship. Even among the few underpowered studies, observed effects are not clearly larger. In fact, the weakest correlation identified in the academic dishonesty search comes from the study with the smallest sample (r = .018, N = 44). Additionally, the correlation between moral neutralization and an outcome was often not the focal analysis, meaning publication likely did not hinge on the relationship's significance. Therefore, I do not hypothesize that the mean effect will be inflated by publication bias. Still, publication bias is assessed because the data is available and it is best practice.

A variety of techniques were used to assess publication bias. Different methods of testing and correcting for publication bias have different strengths and weaknesses, and no method is especially reliable (Inzlicht et al., 2015). Therefore, a variety of methods were used. First, the trim-and-fill method was used. This method generates a funnel plot of observed effects on the X-axis and standard error on the Y-axis, tests it for asymmetry, then imputes studies to make the plot symmetrical (Duval & Tweedie, 2000). The resulting mean and distribution of effects is interpreted as the publication-bias corrected mean and distribution. However, the results must be interpreted with caution because the method often underestimates the degree of publication bias, generates incorrect confidence intervals, and occasionally identifies bias when none is present (Terrin et al., 2003; Peters et al., 2007; Moreno et al., 2009). Therefore, it should not be used on its own.

Next, Egger's regression test, the precision effect test (PET), and the precision effect estimate with standard error (PEESE) were computed. These methods of analyzing publication bias are also based on the funnel plot. Egger's regression test (Egger et al., 1997) can be thought of as a weighted least squares regression where effect size is predicted by the weighted standard error. A significant slope effect (coefficient for the standard error term) is interpreted as evidence of funnel plot asymmetry and publication bias. PET goes one step further and interprets the intercept of this regression as an effect size when standard error is 0. In other words, the intercept can be interpreted as the effect size in a perfectly precise study that is unaffected by publication bias (Stanely, 2005). Simulation studies suggest that PET overcorrects for publication bias when the true correlation is not 0. Given that there is strong theoretical and empirical reason to believe that moral neutralization has non-zero correlations with the outcomes included in the analyses, PET will likely overestimate the impact of publication bias. PEESE considers this flaw by using the squared standard error (i.e., the variance) as the predictor instead of the standard error. These methods are often combined and referred to as PET-PEESE. The PET-PEESE approach begins by estimating PET, and if the observed mean effect is significantly different from 0, PEESE is also estimated as a more conservative correction.

Regression methods for correcting for publication bias (Egger's, PET, PEESE) tend to outperform the trim-and-fill method in simulation studies (Moreno et al., 2009; Rücker et al., 2011). However, they are not without their flaws. First, seeing as an effect with a standard error of 0 cannot exist, interpreting the intercept when conducting PET-PEESE necessarily involves extrapolating beyond the range of observed values used to fit the model. Second, methods of detecting publication bias based on funnel plots can be misleading when between-study heterogeneity is high (Carter & McCullough, 2014; Terrin et al., 2003). One approach to addressing this issue is to conduct the analyses within each moderator category to control for the heterogeneity associated with those moderators. However this was not done in the present study because 1. Many of the examined moderators were not hypothesized to account for substantial heterogeneity, and 2. if publication bias is present, it is likely to be weak. Therefore, methods of detection must include as many effects as possible to have adequate power to detect the weak effect. One final limitation of regression-based approaches is that the results can be influenced by small-study effects other than publication bias. For example, if a literature contains both experimental and correlational studies, it could be that the experimental studies tend to have smaller samples and produce larger effects, on average, due to the controlled laboratory setting. In such a case, PET-PEESE would correct for those small study effects. Therefore, PET-PEESE can be used most appropriately when the only sources of between-study heterogeneity is the result of sampling error and publication bias.

Methods not based on the funnel plot were also used to test for publication bias. The test for excess significance (Ioannidis & Trikalinos, 2007) compares the number of observed significant effects to the expected number of significant effects given the average sample size. If there are many more significant findings than expected, publication bias may be affecting the results. Finally, cumulative meta-analysis (McDaniel, 2009) was also conducted to check for publication bias. In a cumulative meta-analysis, studies are sorted by precision (i.e., standard error or sample size). Beginning with the most precise effect, iterative meta-analyses are conducted, with the next most precise effect being added after each iteration. The mean effect and/or heterogeneity is recorded during each iteration, and the results can be checked for drift. If certain types of effects from smaller, less precise studies are being suppressed, the less precise studies added to the cumulative meta-analysis in the later iterations will produce drift.

Results

The analyses yielded moderate to large correlations between moral neutralization and the six outcome variables (see Table 7). In the primary analyses, the magnitude of the corrected correlations ranged from $\bar{\rho}$ = -.15 (95% CI [-.28, -.03]) with OCBs to $\bar{\rho}$ = .49 (95% CI [.39, .59]) with unethical pro-organizational behaviors. Unethical pro-organization behaviors $\bar{\rho}$ = .49 (95%

CI [.39, .59]), turnover intentions $\bar{\rho}$ = .38 (95% CI [.22, .54]), CWBs $\bar{\rho}$ = .45 (95% CI [.42, .49]), and academic dishonesty $\bar{\rho}$ = .42 (95% CI [.36, .48]) were each strongly correlated with moral neutralization. OCBs $\bar{\rho}$ = -.15 (95% CI [-.28, -.03]) and job satisfaction $\bar{\rho}$ = -.22 (95% CI [-.59, .15]) were moderately correlated with moral neutralization.

For all relationships, heterogeneity was notably high (see 80% credibility intervals and I^2 values in Tables 7-10). I^2 ranged from 50.15% to 93.74%. The next lowest I^2 value was 77.19%. Even in the moderator analyses, heterogeneity was substantial. Although most relationships were consistently in the same direction, the large credibility intervals suggest that the magnitude of effects varied substantially across studies. For example, the 80% credibility interval around the mean correlation between moral neutralization and CWBs was [.27, .63]. These high levels of heterogeneity may be indicative of moderators (Schmidt & Hunter, 1977).

Due to the large number of studies examining CWBs and academic dishonesty, we were able to examine several moderators (see Tables 7-10). However, despite substantial heterogeneity, type of measure, theory the measure was based on, type of CWB, content matching, and CWB specificity did not significantly moderate the moral neutralization and CWB correlation. Although most trends were in the suspected directions (e.g., the Moore et al., 2012 and neutralization measures were more strongly correlated with CWBs than other measures; moral neutralization measures designed to predict cheating were more strongly correlated with academic dishonesty), confidence and credibility intervals overlapped. Notably, despite overlapping confidence intervals, whether moral neutralization measures were matched to the outcome variable or not did substantially moderate the strength of the academic dishonesty relationship (matched $\bar{\rho}$ = .45, unmatched $\bar{\rho}$ = .35), but it did not appear to moderate the CWB relationship (matched $\bar{\rho}$ = .43, unmatched $\bar{\rho}$ = .46).

Several publication bias analyses were conducted on the academic dishonesty and moral neutralization relationship. First, a funnel plot was generated and the trim-and-fill method was used. However, the funnel plot was not asymmetrical, so no new studies were imputed (Figure 8). Next, Egger's regression test was computed, but it was not close to significant. (z = .45, p-value = .65). Next, PET was estimated, and because the PET-estimated effect size was significantly greater than 0 (r = .518), PEESE was also estimated (r = .442; see Figure 9). Methods of testing for publication bias not based on the funnel plot were also used. The test of excess significance found no signs of bias (*p-value* = .36). The number of observed significant findings was 36 out of 38, and the expected number was about 35 (Table 11). The results of the cumulative meta-analysis can be seen in Figure 10.

The aforementioned publication bias analyses were repeated for the moral neutralization and CWB relationship. Again, the trim-and-fill method did not result in any imputed studies (see Figure 11). Additionally, Egger's regression test of asymmetry was not significant (z = 0.37, p = .71). PET and PEESE corrected the correlation upward, which is the opposite of what is to be expected in the typical publication bias situation (Figure 12). It should be noted that the corrections may be inaccurate due to the substantial heterogeneity. The test of excess significance did not produce significant evidence of publication bias. 76 of 77 findings were significant. About 74 were expected to be significant (Table 12). The cumulative meta-analysis also did not show signs of substantial publication bias (Figure 13), although the smallest studies appeared to produce small negative drift, which is consistent with the results of the regression-based methods..

Discussion

The hypotheses related to main effects were supported. Moral neutralization was correlated with all outcomes. Giganc and Szodorai (2016) suggested that correlations of .10, .20, and .30 are relatively small, medium, and large effects, respectively within the context of individual differences research. With those guidelines in mind, it is clear that moral neutralization is strongly correlated with important outcomes. Other scholars have provided similar guidelines for interpreting effect sizes (Funder & Ozer, 2019).

The moderators did not have a statistically significant effect on the mean effect size, though trends were mostly in the expected direction. As expected, type of CWB, CWB specificity, and publication status did not substantially moderate the strength of the correlation. However, I hypothesized that the measure used and whether measure content was matched to the outcome would moderate the moral neutralization and CWB correlation, but the results did not support these hypotheses. The trend was in the predicted direction in the academic dishonesty analyses (measures that asked about moral neutralization related to academic dishonesty, such as Shu et al. (2011), were more strong correlated with academic dishonesty than other moral neutralization measures, though confidence intervals overlapped), but there was a very small difference in the opposite direction in the CWB analyses. This is surprising, seeing as there is evidence that matching the content of a moral neutralization assessment to the target outcome variable can improve prediction (Maruana & Copes, 2005; Study 1). It may be that CWB is a broad outcome, so matching assessment content is difficult and less effective. On the other hand, academic dishonesty is a more specific outcome, so matching the content resulted in a larger difference. Another potential explanation for these conflicting results could be that the domain general measures used in CWB studies are better assessments than the contextualized measures. For example, Moore et al.'s general measure has been extensively validated. The same cannot be said for many of the contextualized measures used. Therefore, the effects of matching content and using stronger measures may be as hypothesized, but because they are both present in the same studies, their effects may be hidden. Another potential explanation for the insignificant findings is that the number of studies in each category is too small, resulting in large standard errors. For example, there were only ten studies in the not-matched category in the academic dishonesty analyses.

Importantly, theory (moral disengagement versus neutralization) did not significantly or substantially moderate the strength of the relationship between moral neutralization and CWB or academic dishonesty. The relationship between key outcome variables appears to be roughly equivalent regardless of whether or not measures are based on techniques of neutralization and academic dishonesty. This is consistent with the previous findings that both are very similar to each other, and that differences may be superficial.

Heterogeneity was high in all analyses, including within each moderator. This suggests that there may be additional moderators not included in the analysis. Although most of the moderators included in the present analysis did not significantly influence the moral neutralization and counterproductivity relationship, previous studies have reported an effect for other types of moderators. For example, Barsky (2011) found that participating in goal setting significantly moderated the relationship between moral justification (a mechanism of moral disengagement) and unethical behavior at work such that participation resulted in a weaker relationship. In another example, Knoll et al. (2016) found that moral neutralization was associated with unethical behavior in weak situations, but not in strong situations. See Newman et al. (2019) for a review of moral neutralization moderators in the workplace. These moderators and more may be producing the large degree of heterogeneity across observed effects, but their presence is seldom reported, so they could not be included in the present meta-analysis.

Publication bias analyses were conducted for the moral neutralization and CWB and academic dishonesty meta-analyses. The publication bias analyses were limited to those two relationships because they had a sufficient number of studies. In both cases, there was not strong evidence of publication bias. If anything, bias corrections would result in a stronger mean correlation, not a weaker correlation as is typically found when estimating the impact of publication bias. Although there is not evidence of publication bias, the results of the methods used should be interpreted with caution. The results may be affected by the extreme heterogeneity present in the analyses.

The results of the meta-analyses provide strong evidence that moral neutralization is associated with important outcomes, regardless of whether measures are based on neutralization theory or moral disengagement theory. However, moderating factors that impact the strength of the association between moral neutralization and counterproductivity remain unclear. In workplace settings, the utility of moral neutralization measurement depends on whether measures of the construct retain criterion-related validity: 1. After controlling for other variables used to predict counterproductivity and 2. In high stakes situations where respondents may be tempted to lie to make themselves appear more desirable. In the remaining studies, I test the incremental validity and the impact of faking on the criterion-related validity of moral neutralization.

Study 3: Incremental Validity

The results of study 2 are clear. Moral neutralization is correlated with important workplace outcomes. However, despite the apparent importance of moral neutralization, very little work has been done on the incremental validity of the construct. That is, it is not clear whether or not moral neutralization is associated with important outcomes after controlling for other common predictors, such as personality traits and general cognitive ability. What little work has been done has focused on either neutralization or moral disengagement exclusively, meaning the measures of moral neutralization may have been deficient. One study found that contextualized measures of neutralization incrementally predicted CWBs, job performance, and GPA over and above the Big Five personality traits and cognitive ability (Lee & Kuncel, 2020). Ogunfowora et al. (2021) used a meta-analytic dataset to test the incremental validity of moral disengagement over “dark” personality traits (narcissism, psychopathy, machiavellianism, and psychological entitlement). They found that moral disengagement predicted workplace misconduct, but not OCBs, after controlling for dark traits. However, some of the effect sizes in the meta-analytic correlation matrix used to conduct the analyses were based on only a few studies, which threatens the external validity of the findings.

All previous research on the incremental validity of moral neutralization fails to use a composite of moral disengagement and neutralization, and only includes a small subset of relevant predictors to control for (e.g., Big Five or Dark Triad as opposed to both simultaneously). In the present study, I seek to determine the extent to which moral neutralization is associated with CWBs and academic dishonesty after controlling for several other common individual difference predictors of deviance. By including several of the most relevant individual difference predictor variables, the test of incremental validity will be especially rigorous.

Among the most common predictors of CWBs and academic dishonesty are the Big Five personality traits, with conscientiousness exhibiting the strongest relationship ($\rho = -.25$ and $\rho = -.24$, respectively; Lee et al., 2020; Ones et al., 1993). Another set of personality variables that are frequently used to predict deviance is the Dark Triad. The Dark Triad consists of machiavellianism, narcissism, and psychopathy. According to a meta-analysis, the mean reliability-corrected correlations between the Dark Triad and CWBs range from .07 to .43 (O'Boyle et al., 2012). I failed to find meta-analytic estimates of the relationships between Machiavellianism and narcissism with academic dishonesty, but Lee et al. (2020) estimated that the reliability-corrected correlation between academic dishonesty and psychopathy to be .40.

Although not studied as frequently, impulsivity is another trait that is especially relevant to the prediction of academic dishonesty ($\rho = .39$; Lee et al., 2020). In theory, the most relevant personality traits affect the likelihood of cheating by influencing the intention to engage in cheating behaviors (Whitley, 1998). However, Lee et al. (2020) suggest that many cheating behaviors (such as peeking at another student's exam) are spontaneous acts that do not involve prior intent. Therefore, impulsivity is an especially important trait to consider when evaluating the personality correlates of academic dishonesty. Similar correlations are observed in studies of impulsivity and CWBs (e.g., Henle, 2005).

Another popular predictor of counterproductivity is cognitive ability. Indeed, cognitive ability is among the most thoroughly studied variables in all of psychology. Researchers have gone so far as to say "g is to psychology what carbon is to chemistry" (Ree & Earles, 1993).

Therefore, it is unsurprising that it is a frequently studied predictor of CWBs and academic dishonesty (Cuadrado et al., 2020; Dilchert et al., 2007), though the observed relationships tend to be weaker than the correlations between cognitive ability and other metrics of performance (e.g., Schmidt & Hunter, 1998).

Because the validity of common predictors such as cognitive ability and the aforementioned personality variables for predicting CWBs and academic dishonesty has already been well established in the literature, the utility of moral neutralization as a predictive construct depends on whether or not it can account for variance in important outcomes beyond that already accounted for by these variables. Ogunfowora (2021) conducted a meta-analysis of moral disengagement correlates and found strong relationships with several of the personality traits described above (see Table 13). Some of these correlations are very strong. For example, the unreliability-corrected correlation between moral neutralization and psychopathy was .76. Due to the strong overlap between these traits, it is reasonable to question whether moral neutralization would account for additional variance in CWBs and academic dishonesty. However, previous research suggests that moral neutralization does incrementally predict counterproductivity after controlling for at least a subset of these traits (Lee et al., 2020; Ogunfowora et al., 2021). Additionally, moral neutralization, with its strong emphasis on rationalization, is conceptually distinct from the other individual difference variables described above. Therefore, I hypothesize:

Hypothesis 1: Moral neutralization will account for variance in self-reported CWBs after controlling for personality, cognitive ability, and demographics.

Hypothesis 2: Moral neutralization will account for variance in self-reported academic dishonesty after controlling for personality, cognitive ability, and demographics.

As discussed previously, a small amount of research has examined the individual difference correlates of moral neutralization. Lee and Kuncel (2020) found that correlations between contextualized neutralization measures and DeYoung et al.'s (2007) big five aspects were weak to moderate, with aspects of conscientiousness and agreeableness exhibiting the strongest correlations. Rengifo and Laham (2022) found much stronger correlations between the propensity to morally disengagement and aspects of the big five. Some correlations were much stronger than is typically found in personality research. For example, the correlation between moral disengagement and the politeness aspect of agreeableness was -.74. However, when moral disengagement was measured via Shu et al.,'s (2011) academic dishonesty-related measure, correlations were weaker (the correlation with politeness dropped to -.32). It may be that personality correlates are stronger when moral neutralization is measured with a relatively decontextualized measure (e.g., Moore et al., 2012). Contextualized measures may be contaminated with situation-specific variance not associated with broad personality traits. Therefore, I hypothesize that:

Hypothesis 3a: The Moore et al. (2012) measure will be more strongly correlated with the big five and the dark triad traits than the contextualized measures.

Additionally, the general factor extracted from Model 4 in Study 1 should be an estimate of moral neutralization after removing the specific variance associated with the measures. Therefore:

Hypothesis 3b: The estimated general moral neutralization factor will be more strongly correlated with the big five and the dark triad traits than scores on the measures.

To my knowledge, no study has explored the individual difference correlates of moral neutralization broadly, instead focusing exclusively on moral disengagement or neutralization. I take this time to investigate the pattern of relationships between a moral neutralization and individual differences.

Exploratory Research Question 1: What is the pattern of correlations between the moral neutralization composite and individual differences?

The inclusion of such a wide range of variables provides an opportunity to gain a better understanding of the factor scores extracted during Study 1. In addition to investigating the relationship between the general moral neutralization factor and personality traits, correlations between study variables and specific factors will also be investigated. In theory, the specific factors account for measure-specific variance unrelated to the broad, moral neutralization construct that accounts for variance in scores across measures. Therefore, I do not expect many strong correlations between personality and the specific factor scores. Still, specific factor score correlations are investigated to better understand the meaning for the Model 4 factors.

Exploratory Research Question 2: What is the pattern of correlations between the specific factors and individual differences?

Methods

Sample

The sample used in this study is the same as the sample described in Study 1. Key variables were measured during each wave of data collection. The total sample consists of 1,443 respondents. 375 of the respondents were recruited on Amazon Mechanical Turk, 328 came from the University of Minnesota's REP system, and 738 were recruited via Prolific.

Materials

Moral Neutralization

One shortcoming of most previous research is the use of only neutralization or moral disengagement measures. Unfortunately, only one measure has been developed to measure moral neutralization directly, and it is meant to only be used on child samples (Ribeaud & Eisner, 2010). To remedy this issue, I included measures of both neutralization and moral disengagement, and then aggregated the results to form a composite measure of moral neutralization. Measures of neutralization and moral disengagement tend to be short and contextualized (they ask about the tendency to invoke rationalizing attitudes in certain contexts, such as when stealing or when at work), so multiple measures are necessary to measure the general moral neutralization construct. I included two self-report measures of moral disengagement consisting of 29 total items (Moore et al., 2012, $\alpha = .83$; Swan et al., 2017, $\alpha = .87$), and three self-report measures of neutralization consisting of 23 total items (Agnew & Peters, 1986, $\alpha = .88$; Haines et al., 1986, $\alpha = .97$; Thurman, 1984, $\alpha = .87$). These five measures are the same as those used in Study 1. Scores were standardized and averaged to form an overall measure of moral neutralization. Because each measure was weighted equally despite differing lengths, reliability could not be estimated by simply computing the Cronbach's alpha of the 42-item test composite (though doing so yields a value of $\alpha = .94$). Instead, the reliability of the composites were estimated via the method described in Nunnally & Bernstein (1994). The reliability of the composite was .97. This composite of neutralization and moral disengagement measures was the primary measure of moral neutralization used in the analysis. However, the general factor extracted from Model 4 in Study 1 was used as an alternative method of measuring moral neutralization.

The Big Five

The Big Five personality traits, agreeableness (α = .89), conscientiousness (α = .91), extraversion (α = .93), openness (α = .88), and neuroticism (α = .94), were measured using the Big Five Aspect Scales (DeYoung et al., 2007; BFAS). The BFAS consists of ten subscales, each with ten items. Scores on each of the Big Five traits are formed by averaging the two subscales associated with the relevant Big Five trait. Conscientiousness is estimated by averaging scores on the orderliness (α = .84) and industriousness (α = .91) scales, agreeableness is estimated by averaging scores on the politeness (α = .76) and compassion (α = .91) scales, neuroticism is estimated by averaging scores on the withdrawal (α = .91) and volatility (α = .90) scales, openness to experience is estimated by averaging scores on the openness (α = .84) and intellect (α = .88) scales, and extraversion is estimated by averaging scores on the assertiveness (α = .91) and enthusiasm (α = .90) scales. In total, the BFAS contains 100 Likert scale items.

The Dark Triad

The Dark Triad, Machiavellianism (α = .85), narcissism (α = .80), and psychopathy (α = .77), were measured via the Short Dark Triad (Jones & Paulhus, 2014; SD3). The SD3 consists of 27 items (nine items for each Dark Triad trait).

Impulsivity

Impulsivity (α = .84) was measured via the 13-item abbreviated version of the Barratt Impulsiveness Scale (ABIS; Paula et al., 2020). An example item on the IBIS is “I do things without thinking”. Responses to each item range from 1 (Rarely/ Never) to 4 (Always / Almost Always). Although the ABIS is fairly new, preliminary research suggests that it can provide reliable and valid information.

Cognitive Ability

Cognitive ability (α = .75) was measured by the 16-item International Cognitive Ability Resource test (International Cognitive Ability Resource Team, 2014; ICAR). This version of the ICAR includes four verbal reasoning items, four letter and number series items, four matrix reasoning items, and four three-dimensional rotation items. While collecting data from the Prolific sample, the average time to complete the survey was longer than expected. Therefore, to save time, the ICAR was removed from the survey midway through data collection. In total, 375 MTurk, 328 REP, and 345 Prolific respondents completed the ICAR. Due to the large amount of missing data, (393 missing Prolific responses), analyses were conducted twice. Once with the ICAR included, and once with it removed to increase the sample size.

Demographics

Respondents were asked to report their age and gender (Male = 0, Female = 1, other options excluded from analysis). I controlled for both these two variables and whether or not the respondent was a REP participant, an Amazon Mechanical Turk participant, or a Prolific participant. In the full sample, 678 respondents identified as male. 740 identified as female. Of the REP respondents, 76 identified as male and 245 as female. Of the MTurk respondents, 202 identified as male and 172 as female. Of the Prolific respondents, 400 identified as male and 323 identified as female.

Counterproductive Work Behaviors

To measure CWBs (α = .88), I used the Measure of Workplace Deviance (Bennett & Robinson, 2000) described in Study 1. The Cronbach's alphas of the CWB-I and CWB-O scales were .83 and .84, respectively.

Academic Dishonesty

Academic dishonesty (α = .91) was assessed by McCabe and Trevino's (1993) 12-item self-report measure. This is the same tool described in Study 1.

Infidelity Intentions

Data on infidelity intentions was available for 670 of the Prolific participants. Therefore, I investigate it as an additional criterion variable due to its availability. Infidelity intentions (α = .75) was measured via the ITIS described in Study 1.

Analysis

All analyses were conducted in R. First, correlations between moral neutralization and all study variables were computed in the total sample. Next, regression models were created for each outcome variable. The null models consisted of all the control variables being regressed on the outcome variables. To test the hypotheses, the moral neutralization composite was added to the models and the coefficients of determination (R^2) were compared. An increase in R^2 when predictors are added suggests that the added predictors account for additional variance in the outcome variable, providing support for the hypotheses. The regression analyses were conducted twice. Once with the subset of data containing the cognitive ability assessment, and again with the full sample but without the cognitive ability assessment.

Results

Correlations between the moral neutralization measures, composite, and factor can be seen in Table 14. The correlations between moral neutralization and other predictors and outcomes can be seen in Table 15. The correlations between other predictors with outcomes can be seen in Table 16. Finally, the results of the regression models can be seen in Tables 17-22. The results of the regression analyses did not substantially differ regardless of whether or not cognitive ability was included in the model. Therefore, the larger sample-size models without cognitive ability are reported and discussed here. The results of the analyses using the models with cognitive ability can be viewed in Appendix 7.

In all cases, moral neutralization was a significant predictor of the outcome variable after controlling for all other predictors. For each outcome, adding the moral neutralization composite increased the proportion of variance accounted for by at least 1% (ΔR^2 ranged from .10 (infidelity intentions) to .08 (academic dishonesty)). The addition of the moral neutralization factor instead of the composite resulted in similar results, although ΔR^2 was generally slightly smaller (ΔR^2 ranged from .006 (CWB-I) to .073 (academic dishonesty)). In all cases, the increase in R^2 was statistically significant at the $\alpha = .01$ level, and a substantial amount of variance in the outcome was accounted for. These results provide support for the hypotheses which posit that moral neutralization accounts for variability in counterproductive outcomes above and beyond what is accounted for by common individual difference predictors.

The moral neutralization composite was moderately to strongly related to several personality traits. Most notably, substantial relationships were observed for conscientiousness (r = -.34), agreeableness (r = -.46), neuroticism (r = .28), Machiavellianism (r = .51), narcissism (r = .22), and psychopathy (r = .55). At the Big Five aspect level, moral neutralization was most strongly associated with politeness (r = -.48), followed by compassion (r = -.33) and industriousness (r = -.32). The moral neutralization general factor was similarly correlated with the other variables. To address hypothesis 3, correlations between the Big Five, Dark Triad, and individual measures of moral neutralization were examined. Of the measures, Moore et al.'s (2012) measure usually correlated more strongly with the other predictors than the other measures (mean r = .33), followed by Thurman's (1984) measure (mean r = .26). In most cases, Moore et al.'s measure was a significantly stronger correlate of personality than the other measure. For example, the Moore measure was significantly most strongly correlated with Agreeableness (Moore r = -.48, Swan r = -.37, p = < .001). However, there were exceptions (see Table 15). Similarly, the moral neutralization factor tended to be more strongly correlated with the Big Five and Dark Triad than the individual measures (mean r = .34). The specific factors were largely uncorrelated with the study variables (Table 23)

Discussion

The results provide support for the hypotheses. I hypothesized that moral neutralization would account for variance in self-reported CWBs (Hypothesis 1) and academic dishonesty (Hypothesis 2) after controlling for personality, cognitive ability, and demographics. In support of these hypotheses, moral neutralization predicted CWBs and academic dishonesty even after controlling for common individual difference variables. Additionally, in most cases, the increase in R^2 was quite substantial, indicating that moral neutralization accounts for variance in the outcomes not associated with the other predictor variables. Findings were similar when a measure of infidelity intentions was used as the criterion variable. Therefore, it stands to reason that moral neutralization may be useful in applied settings where predicting deviant behavior is a goal of decision makers. For example, some of the most popular predictors of CWBs used in selection are the Big Five personality traits (in various forms). Because moral neutralization predicts deviance beyond those common predictors, it may be useful in selection contexts.

Although the ΔR^2 was substantial in all models, the improved prediction of academic dishonesty is especially noteworthy. The relationship between the other predictors and academic dishonesty was relatively weak compared to CWB (baseline $R^2 = .209$ compared to .279 in the baseline CWB models). However, introducing the moral neutralization composite to the model resulted in a large increase in R^2 ($\Delta R^2 = .080$). This value is almost twice as much as the also large ΔR^2 that resulted from adding the moral neutralization composite to the CWB models. This large improvement to the academic dishonesty model could be the result of including the Haines et al. (1986) measure in the composite. The Haines et al. measure includes items such as “Jack should not be blamed for cheating if the instructor doesn’t seem to care” and is designed specifically to predict cheating behavior. However, the general moral neutralization factor is meant to be an estimate of moral neutralization with the contextualized measure variance removed, and including the factor in the academic dishonesty model resulted in a similarly strong ΔR^2. It may simply be the case that moral neutralization is an especially strong predictor of academic dishonesty, even after controlling for other individual differences.

Results differed depending on the type of CWB used as the outcome variable. Although adding moral neutralization to the model resulted in improved prediction of both CWB-Is (ΔR^2 = .015) and CWB-Os (ΔR^2 = 0.40), the gain was larger in the CWB-O models. It may be that it is easier for people to invoke rationalizing attitudes to justify antisocial behavior targeted towards faceless institutions that hold power over them than it is when the target of the antisocial behavior is another person.

Where the participant was recruited from (Prolific, MTurk, or REP) was a significant predictor of the outcomes, especially in the CWB models. This is consistent with the observation that MTurk and Prolific respondents reported engaging in more CWBs, on average, than REP respondents (See Appendix 6 for descriptive statistics). However, standardized mean differences were modest (CWB *d* = .28 [.15, .42] for Prolific; *d* = .36, [.21, .50]) for MTurk. These differences may be due to the fact that the REP participants tend to have less work experience so have had fewer opportunities to engage in CWBs. The scale used to measure CWBs (Bennett & Robinson, 2000) uses a frequency response option, so less work experience and weekly time at work could result in a lower score. However, age was also included in the regression models, which may capture the variance in CWBs associated with experience. Regardless, these differences raised the concern that the relationship between the predictors and the criteria may differ depending on the subsample, so the regression analyses were conducted again within each group. The tables containing the regression coefficients and R^2 values from these analyses can be seen in Appendix 7. The ΔR^2 tended to be lower in the Prolific models, but the overall patterns of results were similar. Therefore, the total sample models are the primary focus of discussion.

A review of the moral neutralization measure correlates reveals that Hypothesis 3 was also supported. The domain-general Moore et al. (2012) measure of moral neutralization and the moral neutralization factor extracted during Study 1 both tended to be more strongly correlated with other personality trait measures. This provides further support for the proposition that contextualized measures may contain situation-specific variance not associated with broad personality traits, and that the general factor extracted from Model 4 in Study 1 is an estimate of moral neutralization that is relatively unaffected by contextualized items. However, correlations with the general factor must be interpreted with caution because factor scores can be unstable. Although the results of Study 1 suggest that the general factor is relatively stable (Guttman's Indeterminacy Index = .824), correlations with the variables included in this study could vary substantially with an alternative set of factor scores. An indeterminacy index of .824 means that an alternative set of factor scores could be generated that correlate only .824 with the current set of factor scores. .824 is a strong correlation, but such a relationship allows for variability in relationships with other variables.

An examination of the personality correlates of moral neutralization reveals an interesting pattern. The personality correlates of moral neutralization were similar to those of integrity tests. Conscientiousness, agreeableness, and neuroticism were the strongest big five correlates of moral neutralization, just as they are for overt integrity tests (Berry et al., 2007). This is, perhaps, unsurprising given the conceptual similarity between moral neutralization and integrity. The similarity between moral neutralization and integrity is revisited in Study 4.

All three dark triad traits were also associated with moral neutralization. The traits most strongly associated with moral neutralization are also those most strongly associated with deviance and unethical behavior (Ellen et al., 2021). Despite the moderate to strong correlations with other predictors, moral neutralization still predicted counterproductive behavior after controlling for these traits. These results provide support for the convergent and divergent validity of moral neutralization.

The pattern of correlations between the Model 4 specific factors associated with each moral neutralization measure and other study variables was about as expected (Table 23). Correlations between the specific factors and their associated measures were strong, but the correlation between the Moore et al. (2012) measure and specific factor was relatively weaker. Moore et al.'s measure is domain-general, whereas the other measures assess moral neutralization related to a specific types of behavior, such as academic dishonesty or driving. Given that the specific factors consist of variance unrelated to the general moral neutralization factor, a weaker correlation with the domain-general moral disengagement measure is consistent with theory. Correlations between specific factors and personality traits were near zero, with only a few exceptions. For example, the Moore et al. factor was correlated with politeness and compassion. However, even when correlations were present, they were typically weaker than the correlations with the general moral neutralization factor. These findings provide support for the idea that the general factor is an estimate of overall moral neutralization and the specific factors are made up of context-specific variance largely unrelated to stable, domain-general individual differences. Note that some of the specific factors are very unstable. The Guttman Indeterminacy Indices for the specific factors range from .422 (the Thurman factor) to .868 (the Haines et al. factor; see Table 4). Therefore, correlations with these factors must be interpreted only with extreme caution.

Moral neutralization appears to be associated with important outcomes, even after controlling for other common predictors. However, before moral neutralization assessments are used in applied settings, more research is needed. Given that moral neutralization and integrity have similar correlates, the differences between the constructs must be explored. Additionally, it is not clear that the results of the present study would generalize to applied settings, where respondents may be tempted to fake their responses such that they appear more desirable (e.g., when applying for a job). Therefore, it is important to assess the extent to which moral neutralization assessments can be faked.

Study 4: Directed Faking

When considering the construct validity of moral neutralization, the Big Five correlations estimated in Study 3 are especially notable (Table 15). Moral neutralization was correlated with conscientiousness ($r = -.34$), agreeableness ($r = -.46$) and neuroticism ($r = .28$). Previous research has revealed similar relationships. For example, in a meta-analysis, Ogunfowora et al. (2021) found a correlation of -.38 between moral neutralization and conscientiousness ($r = -.30$ before corrections). These correlations with the Big Five are similar to those found in studies of integrity tests (Berry et al., 2007). Integrity can be conceptualized as a compound trait consisting of conscientiousness, agreeableness, and neuroticism (Ones, 1993). The Big Five correlations suggest possible overlap between measures of moral neutralization and integrity.

Integrity tests can be classified into two broad categories: personality-based and overt integrity tests (Sackett et al., 1989). Personality-based integrity tests attempt to measure integrity obliquely by assessing a variety of correlated personality dimensions, such as conscientiousness and trustworthiness. Overt integrity tests take a more direct approach by assessing attitudes towards dishonesty, and sometimes even asking about previous counterproductive behaviors. Both types of integrity tests tend to be highly correlated, and both predict organizational outcomes effectively; however, overt integrity tests tend to be a stronger predictor of CWBs (Ones et al., 1993). Integrity tests predict CWB even after controlling for other common predictors, such as the Big Five personality traits (Marcus et al., 2013).

Although personality and integrity tests are powerful predictors of CWB and are widely used by practitioners, some concerns remain because responses to personality and integrity tests can be faked by respondents seeking to make themselves look better than is accurate. One of the most common methods of assessing the fakability of an assessment is to simply direct research participants to distort their responses in a certain way, then compare those responses to truthful responses. If the responses in the directed faking condition differ, the effect size describes the potential impact of distorted responding. Viswesvaran et al. (1999) conducted a meta-analysis to evaluate the fakability of Big Five assessments and found that measures of all five traits were highly fakable. Standardized mean differences ranged from .47 to .93 standard deviations in studies where respondents were instructed to fake "good."

Similarly, integrity tests also appear to be quite fakable. In a meta-analysis, overt integrity tests were more fakable than personality-based tests, with an effect size of .93 (compared to .38) standard deviations in fake good conditions (Alliger & Dwight, 2000). However, this effect may be inflated due to methodological issues with the meta-analysis (Berry et al., 2007). Still, most primary studies find large faking effects (e.g., Hurtz & Alliger, 2002; Ryan & Sackett, 1987). As a result, scholars have searched for methods that can be used to detect or suppress distorted responding. For example, researchers have attempted to reduce the fakability of integrity tests by using a forced choice response format (Jackson et al., 2000). Response latency can also be used to identify fakers (Dwight & Alliger, 1997; Holden, 1995). Despite their weaker predictive power, personality-based integrity tests are sometimes used because they are more difficult to fake than the more predictive overt personality tests (Veris Benchmarks, n.d.). An alternative approach is to use a similar but different predictor that is less prone to being faked.

Moral neutralization and overt integrity tests are strong predictors of counterproductive behavior, and estimates of their correlations with CWBs are similar (Study 2; Ones et al., 1993). Additionally, as mentioned previously, both constructs' relationships with personality traits are similar. The strongest Big Five correlates of both integrity and moral neutralization are conscientiousness, agreeableness, and neuroticism, and they both incrementally predict counterproductivity after controlling for those traits (Ones, 1993). Also, some integrity test items are similar to the content found in moral neutralization measures (for example, 'It's O.K. to lie about the past to help get a job if you will be very honest after you're hired" [Ryan & Sackett, 1987]). Some integrity items involve rationalization in a manner similar to moral neutralization, though typically not to the same degree as moral neutralization items.

It is clear that moral neutralization measures are similar to integrity tests, and that they are strong predictors of counterproductivity. What is less clear is the extent to which respondents can fake their responses to moral neutralization items to appear more desirable. Virtually all measures of moral neutralization are self-report, and items intuitively appear to be highly fakeable. However, there are theoretical reasons to believe that faking a self-report moral neutralization assessment is more difficult than it appears. A core tenet of neutralization and moral disengagement theory is that people invoke rationalizing attitudes, and as a result they actually believe that counterproductive behaviors are justified. Moral neutralization involves “the cognitive reconstruction of inhumane conduct into a benign or worthy [conduct.]” (Bandura, 1999, p. 193) If a respondent is high in moral neutralization and if a justification is present in the item stem, the perceived ideal response could be to actually engage in the counterproductive behavior. Therefore, if a respondent is attempting to fake their responses such that they appear “good,” the respondent could still endorse the item if they have a tendency to believe that the rationalization strategy justifies the behavior. Some items on integrity tests also include rationalization components, such as the item from Ryan and Sackett (1987) mentioned previously, but rationalization is not typically a core aspect of the integrity construct. As a result, integrity tests lack the theoretical protection against faking that is present for moral neutralization measures. Moral neutralization assessments may be suitable alternatives to integrity tests that are relatively resistant to distorted responding, even more so than personality-based integrity tests. The present study begins to consider this possibility by comparing the impact of directed faking on responses to integrity tests and moral neutralization. To assess the impact of faking on moral neutralization measures compared to integrity tests, I conducted a within-subjects directed faking study. My hypotheses are as follows:

Hypothesis 1a: *Directed faking compared to truthful responding will produce a stronger standardized mean difference for overt integrity attitudes than for moral neutralization.*

Hypothesis 1b: *Directed faking compared to truthful responding will produce a stronger standardized mean difference for overt integrity admissions than for moral neutralization.*

Hypothesis 1c: *Directed faking compared to truthful responding will produce a stronger standardized mean difference for personality-based integrity than for moral neutralization.*

Hypothesis 2a: *In the truthful responding condition, moral neutralization will account for variance in counterproductivity after controlling for the Big Five.*

Hypothesis 2b: *In the truthful responding condition, integrity attitudes will account for variance in counterproductivity after controlling for the Big Five.*

Hypothesis 3a: *In the directed faking condition, moral neutralization will still account for substantial variance in counterproductivity after controlling for the Big Five.*

Hypothesis 3b: *In the directed faking condition, integrity will account for less variance in productivity after controlling for the Big Five than moral neutralization.*

Despite their similarities, moral neutralization and integrity are distinct constructs. Moral neutralization has a stronger rationalization component that enables deviant actors to behave unethically without feeling guilty. Therefore, moral neutralization should correlate with counterproductivity even after controlling for the similar integrity construct.

Hypothesis 4a: *Moral neutralization will account for variance in self-reported CWBs after controlling for overt integrity attitudes.*

Hypothesis 4b: *Moral neutralization will account for variance in self-reported CWBs after controlling for overt integrity admissions.*

Hypothesis 4c: *Moral neutralization will account for variance in self-reported CWBs after controlling for personality-based integrity.*

Methods

Participants and Procedure

The sample consists of a subset of the sample used in Studies 1 and 3. The full sample could not be used because not all participants were administered integrity tests. In total, 1,078 respondents completed the overt integrity test (156 from REP, from 187 MTurk, and from 735 Prolific). Only 610 (all from Prolific) completed the personality-based integrity test.

The survey included five broad sections: the truthful responding section, the directed faking section, the personality and abilities section, the outcomes section, and another section containing a decision-making task not included in the current study. The truthful and directed faking sections included the measures of moral neutralization and integrity, and respondents were asked to respond truthfully or to fake their responses. Specifically, in the faking condition, respondents were asked to "please fake your responses as though you are responding to the questions while applying for a job and want to maximize the probability that you will be hired. You DO NOT need to respond truthfully. Respond how you believe the ideal job applicant would respond." In the truthful responding sections, respondents were asked to "Please respond TRUTHFULLY." To ensure that respondents remembered whether or not they were meant to fake or not, at the beginning of each survey block, they were reminded to either respond truthfully or to fake their responses, depending on the section. In the other variables section, respondents were asked about their counterproductive behavior and their personality traits. The order of the survey sections was randomized to control for fatigue effects. However, blocks within a section were randomized for each respondent then administered consecutively. For example, once a respondent was exposed to the faking prompt, they completed each of the faking survey blocks in a random order before moving on to a different section.

Materials

Moral Neutralization

Moral disengagement and neutralization were measured using the same five assessments described in Studies 1 and 3 (Agnew, 1986; Haines et al., 1986; Moore et al., 2012; Swan et a., 2017; Thurman, 1984). Like in Study 3, the measures were standardized and formed into a unit-weighted composite measure of moral neutralization. The moral neutralization measures were administered twice each – once in the directed faking condition and once in the truthful condition. The reliability of the composites were estimated via the method described in Nunnally & Bernstein (1994; see Study 1). The reliability of the composite was .97 in the truthful responding condition and .88 in the directed faking condition.

Overt Integrity

Integrity was measured with the integrity attitudes subscale of the Employee Integrity Index (Ryan & Sackett, 1987). The Employee Integrity Index also includes an admissions subscale. Some studies combine admissions and attitudes subscale scores when estimating the impact of faking (e.g., Hurtz & Alliger, 2002). Both subscales were used in the analyses, but most discussion is focused on the attitudes subscale because it is more similar to moral neutralization measures. The attitudes subscale consists of 52 Likert style items, and Cronbach's alpha was .925 and .958 in the truthful and directed faking conditions, respectively. The admissions subscale consists of 11 items. Cronbach's alpha was .855 in the truthful responding condition and .944 in the faking condition.

Personality-Based Integrity

Personality-based integrity was measured via Veris Benchmark's Trust Profiler assessment. The Trust Profiler consists of 129 questions and is designed to aid employers making hiring decisions by identifying candidates who "work hard, abide by regulations, and in general show more productive, high-integrity behavior (Veris Benchmarks, n. d.)." The Trust Profiler was designed to be resistant to faking. This was accomplished by making it difficult to identify the "best" answer, minimizing a judgmental tone in question stems, and not including admission/confession-based items. Because the Trust Profile is a proprietary assessment, I am not able to provide the items or item-level statistics (including inter-item correlations necessary to compute reliability). However, an example item from the public Veris Benchmarks website is "How much of the time do you: Ignore the rules when they should be ignored?"

Big Five

As in Studies 1 and 3, conscientiousness, agreeableness, neuroticism, openness to experience, and extraversion were measured with the Big Five Aspect Scales (BFAS; DeYoung et al., 2007).

Counterproductive Work Behaviors

CWBs were measured with the Measure of Workplace Deviance (Bennett & Robinson, 2000). This is the same assessment used in Studies 1 and 3. CWB-I, CWB-O, and overall CWB are all included as outcome variables in the analyses.

Academic Dishonesty

To assess another form of counterproductivity, academic dishonesty was measured via the 12-item self-report measure described by McCabe and Trevino (1993).

Infidelity Intentions

Infidelity intentions were measured using the ITIS (Jones et al., 2011) described in Study 1.

Analyses

All analyses were conducted in R. To form the moral neutralization composite variables, scores on moral neutralization measures were converted to Z scores and then averaged. To test Hypothesis 1, I compared the standardized mean difference (Cohen's d) between true integrity and faked integrity to the Cohen's d between mean true moral neutralization and mean faked moral neutralization. Both moral neutralization composites were standardized such that their mean was 0, so rather than compare composite scores, the mean Cohen's d value for the individual moral neutralization measures was used.

To test hypotheses 2 and 3, regression models were fit for each of the counterproductive outcome variables. In the first set of models, only the Big Five personality traits and participant source (Prolific, Amazon Mechanical Turk, or REP) were included as predictors. In the second and third set of models, true moral neutralization or true integrity were added to the Big Five models to evaluate the increase in variance accounted for. An increase in R^2 can be interpreted as evidence in support of hypothesis 2. Finally, in the fourth and fifth set of models, faked moral neutralization or faked integrity were added to the Big Five models. If ΔR^2 is greater in the faked moral neutralization models than in the faked integrity models, hypothesis 3 would be supported. To test hypothesis 4, baseline models containing only one of the true integrity tests were compared to models containing a true integrity test and the true moral neutralization composite. An increase in R^2 when moral neutralization is added to the model means that moral neutralization accounts for variance in the outcomes after controlling for the integrity test. This is interpreted as support for hypothesis 4.

Results

Correlations between all variables included in the analyses can be seen in Table 24, and the values of Cohen's *d* for the measures of moral neutralization and integrity can be seen in Table 25. A positive *d* value indicates that respondents received higher scores while faking. The mean Cohen's *d* of the moral neutralization measures was -.83 (with truthful respondents receiving higher scores, on average), but *d* varied considerably depending on the specific moral neutralization assessment. Moral neutralization *d* ranged from -.62 on Moore et al.'s (2012) moral disengagement scale to -1.22 on Haines et al.'s academic dishonesty neutralization scale. Cohen's *d* for integrity differed depending on the operationalization. While faking their responses, participants scored much higher on the test of overt integrity attitudes than when they were instructed to respond truthfully ($d = 1.24$). However, differences on the admissions test were much smaller ($d = -.56$), and the standardized mean difference for the personality-based integrity tests was smaller-still ($d = 0.31$).The differences in integrity Cohen's d depending on operationalization were all statistically significant at the .05 level.

The results of the regression analyses comparing the incremental validity of true versus faked responses to integrity and moral neutralization tests are summarized in Tables 26 - 28. Consist with the results of Study 3, adding moral neutralization to the model substantially and significantly increased the prediction of all outcomes after controlling for the Big Five, and similar improvements to prediction were observed when each of the different integrity tests were added to the baseline models. In the truthful responding moral neutralization models, ΔR^2 values were .071, .143, and .048 for CWBs, academic dishonesty, and infidelity intentions, respectively. ΔR^2 in the truthful integrity models ranged from .029 (the personality-based integrity test predicting academic dishonesty) to .158 (the overt admissions test predicting CWB).

The incremental validity of the faked integrity and moral neutralization scores over the Big Five (which were *not* faked) was less impressive. None of the integrity tests resulted in statistically significant incremental prediction of infidelity intentions, but moral neutralization did ($\Delta R^2 = 0.13$). Only the overt admission test incrementally predicted CWBs ($\Delta R^2 = 0.005$), and although all of the tests incrementally predicted academic dishonesty, the faked score ΔR^2 was substantially and significantly smaller than ΔR^2 in the truthful responding models. In the models predicting academic dishonesty and infidelity, faked moral neutralization did result in more incremental prediction than the integrity tests (see Tables 27 and 28). However, differences were small in some cases and confidence intervals overlapped substantially .

Table 29 contains the results of the regression that describe the incremental validity of moral neutralization over integrity. Adding moral neutralization to the models resulted in a significant ΔR^2 regardless of which outcome was being predicted or how integrity was measured. ΔR^2 ranged from .011 (predicting CWB over overt integrity attitudes) to .136 (predicting academic dishonesty over personality-based integrity). Moral neutralization yielded statistically significant incremental prediction in all cases, even when overt integrity attitudes and admissions were both included in the baseline model.

Discussion

The results were more nuanced than hypothesized. As hypothesized in hypothesis 1a, the difference between true and faked responses to the overt integrity attitudes test (d = 1.24) was much larger than the difference in scores on the moral neutralization composite (d = -.83). However, differences in integrity admissions (d = -.56) and personality-based integrity (d = .31) were substantially smaller, meaning hypotheses 1b and 1c were not supported. Inspecting the individual moral neutralization scale *d*s reveals that scores on the Haines et al. (1986) academic dishonesty neutralization scale were especially affected by faking, but the same pattern of effects remains even when that scale is removed from the composite (mean d becomes -.74).

To further investigate the impact of faking on the distribution of scores on moral neutralization and integrity tests, histograms for each variable were plotted (Figure 14). In the truthful responding condition, moral neutralization, overt attitudes, and personality-based integrity scores were roughly normally distributed. However, the admissions scale had a strong, positive skew, and the mode was the minimum score of 1.0 (over 16% of respondents). These results suggest a floor effect, which can prevent scores from being lowered in the directed faking condition. If respondents produce a score near the minimum value when responding truthfully, it is not possible for them to substantially lower their score in the directed faking condition. This could explain why the standardized mean difference in admissions scores was lower than expected. If respondents were able to receive lower scores, or if the participants scored higher when responding truthfully, the effect size likely would have been much larger.

In the directed faking condition, moral neutralization and overt integrity attitudes both produced skewed distributions. However, in both cases, there was still a large degree of variability present in the data, and the modal/average response was not the minimum score, indicating that scores are not severely limited by floor/ceiling effects. However, being directed to fake did dramatically increase the number of respondents who received a minimum moral neutralization score and maximum integrity attitudes score, suggesting that faking did not preserve the rank order of scores. Further illustrating this point is the observation that tests are only weakly correlated across the truthful and faked responding conditions, despite strong reliability in both conditions (Table 24). The distribution of scores on the personality-based integrity test also changed in the directed faking condition, but not in the same way as the other distributions. The distribution of faked scores resulted in a relatively flat distribution. The number of respondents who received a strong score on the assessment increased, but so did the number of respondents who received poor scores. Consistent with the design philosophy of the test, faking on the personality-bases integrity test appears to be especially difficult and can often backfire.

Although means and distributions changed in predictable ways in the faking condition (Table 30), some respondents were able to fake desirable responses much more effectively than others. The rank-order of respondents was not consistent across conditions; the correlations between the same tests across conditions were extremely weak (ranging from .08 to .19; see Table 24). Although the floor effects observed in the faking condition may explain some of this change in rank-order, it does not account for all of it. Ranges and standard deviations were fairly stable across conditions. It may be that participants higher in cognitive ability were able to identify opportunities to improve their score more effectively. To test this possibility, I computed correlations between the ICAR and differences in moral neutralization and integrity test scores across conditions. In all tests, cognitive ability was correlated with improving one’s score while faking, but effects were weak, ranging from r = .10 to .19 (Table 31). It is tempting to suggest that correlations are weak because respondents high in cognitive ability may have lower moral neutralization scores to begin with, but moral neutralization was uncorrelated with cognitive ability. The mechanisms by which some respondents are able to improve their scores more than others remain unclear. Future research should attempt to address this gap by administering cognitive interviews to respondents who are faking the test to learn more about the response processes that effective fakers employ compared to less effective fakers.

Although an examination of the standardized mean differences and changes to the distribution of scores when respondents fake is informative, to understand the impact of faking on criterion-related validity, we must turn to the regression models. As expected, moral neutralization and integrity did predict each of the outcome variables beyond the Big Five, providing support for hypothesis 2. However, hypothesis 3 received mixed support. Incremental validity was much weaker in the directed faking condition. Only the integrity admissions test resulted in a statistically significant incremental prediction of CWBs (ΔR^2 = .005), but even then, the effect was weak. The admissions scale includes items that could reasonably be present on a CWB scale (e.g., "Have you ever given unauthorized discounts to friends?"), so despite its statistical significance, this ΔR^2 value is quite low, especially compared to the very large ΔR^2 in the truthful responding model (.158). This suggests that the scale's validity is threatened when respondents are directed to fake their responses.

Faked moral neutralization did significantly incrementally predict academic dishonesty and infidelity intentions. Additionally, the ΔR^2 was larger than in the integrity test models. This provides partial support for hypothesis 3a and 3b. Moral neutralization appears to be an especially effective predictor of academic dishonesty, even when respondents are faking their responses ($\Delta R^2 = .028$). However, even this moderate effect pales in comparison to the large amount of incremental prediction in the truthful responding models. It should also be noted that in the truthful responding condition, integrity outperformed moral neutralization in some cases.

Although faked moral neutralization and integrity responses were much less predictive than in the truthful condition, it should be noted that the degree of response distortion present in this study is unlikely in applied settings. Seeing as respondents were explicitly instructed to fake their responses, this can be interpreted as the maximum impact of faking (assuming attentive responding, which was controlled for with attention check items). Even in high-stakes settings where faking is probable, it is unlikely to be maximized. To some degree, respondents will be honest and other strategies meant to limit faking may be employed. Additionally, in this study, respondents were directed to respond truthfully to the personality assessments, but in an applied setting where respondents are tempted to fake their responses, they would likely fake their personality test responses as well. In applied settings, faking is unlikely to hinder the incremental validity of moral neutralization and integrity tests to the same extent as it did in this study.

Moral neutralization was strongly correlated with overt integrity attitudes ($r = -.77$, $-.74$, and $-.63$ with the composites of moral neutralization, neutralization, and moral disengagement, respectively). Due to the conceptual similarity of moral neutralization and integrity, there was some concern that measures of each were tapping into the same construct. The regression models summarized in Table 29 begin to dispel these worries. In support of hypothesis 4, moral neutralization significantly accounted for variance in each of the outcomes after controlling for each measure of integrity. Moral neutralization incrementally predicted the outcomes even when overt integrity attitudes and admissions were both included in the baseline model. The academic dishonesty and integrity models are especially notable; when integrity admissions and moral neutralization are both included in the model, the overt attitudes regression coefficient drops to near 0. This may hint at an answer to the question raised by Prof. Deniz Ones: what is the nature of the unique variance contained in Integrity measures that leads to incremental predictive power beyond the Big Five? The incremental validity of integrity over its component Big Five traits (conscientiousness, agreeableness, and neuroticism) may be the result of respondent's willingness to admit to admit to some amount of deviant behavior (the admission scale) and how easily respondents can rationalize those behaviors (moral neutralization). Moral neutralization assessments may primarily measure that especially predictive rationalization component of integrity.

It should be noted that the degree of incremental validity of moral neutralization over integrity depended on how integrity was measured. For example, moral neutralization only slightly incrementally predicted CWBs over overt integrity attitudes ($\Delta R^2 = .011$), but when integrity was measured using the personality-based integrity test, incremental validity was much greater ($\Delta R^2 = .088$). Integrity is a broad construct, and different tests measure different dimensions of integrity. Wanek et al. (2003) identified 23 different themes measured by integrity tests that load on four components. Only a small subset of those themes were included in these analyses. Ryan and Sackett's (1987) overt personality test measures two of these themes (honesty attitudes and theft admissions), both of which load on the antisocial behavior component. On its face, the Veris Truth Profiler seems to primarily include themes of orderliness and diligence, which load primarily on a different component. The relationship between moral neutralization and integrity appears to depend on the themes measured by a given integrity test. Future research on the relationship between moral neutralization and integrity should examine a wider range of integrity dimensions. Similarly, the certain techniques of neutralization and mechanisms of moral disengagement may be more strongly associated with integrity than others. Future research should explore the different relationships between the facets of moral neutralization and integrity dimensions.

Overall, the results of Study 4 suggest that integrity and moral neutralization scores can both be affected by faking, but moral neutralization retains a small amount of incremental validity when predicting certain outcomes. On the other hand, the integrity tests, even the personality-based integrity tests designed to be resistant to faking, tended to be slightly less able to contribute to the prediction of deviant behavior over the Big Five in a directed faking study..

Summary

The goal of the present dissertation was to gain a deeper understanding of the broad moral neutralization construct. Scholars have studied moral disengagement and neutralization largely independently for decades, despite their similarity. Both constructs describe a tendency to invoke rationalizing attitudes to justify antisocial behavior. In Study 1, I began by investigating the factor structure of moral neutralization measurement. Five different measures intended to measure either moral disengagement or neutralization were included in several confirmatory factor models. Although my hypotheses were not all supported, moral disengagement and neutralization were strongly correlated and a model containing a broad, overall moral neutralization factor fit the data adequately. Additionally, this general factor was related to other variables in a way that is consistent with the moral neutralization nomological network.

The findings of Study 2 reaffirm the similarity of moral disengagement and neutralization by outlining the virtually identical relationships they have with various criteria in a meta-analysis. Indeed, moral neutralization proved to be a strong correlate of CWBs, academic dishonesty, and other important outcomes. Although moral neutralization was consistently and positively correlated with counterproductivity across studies included in the meta-analysis, the extent to which the findings generalize to applied settings was unclear. The vast majority of the studies included were conducted in low stakes settings, and only raw, bivariate correlations were included in the analysis. Studies 3 and 4 aimed to address these shortcomings by investigating the predictive validity of moral neutralization when a myriad of other constructs were also measured and included in the prediction model, and by testing the impact of faking on moral neutralization compared to integrity. After controlling for several other individual difference variables, moral neutralization predicted counterproductive behavior. Additionally, even when respondents were instructed to fake their responses such that they appeared as desirable as possible, moral neutralization predicted counterproductive behavior in some cases, although prediction was severely impaired.

These findings suggest that moral neutralization could be of use in applied settings. Minimizing counterproductive behavior is a core component of strong job performance (Rotundo & Sackett, 2002), so the strong relationship between moral neutralization and CWBs may be of particular interest to employers. The corrected correlation of $\bar{\rho} = .45$ rivals the strongest individual difference correlates of job performance (Sackett et al., 2021; Schmidt & Hunter, 1998). This effect may even be an underestimate of the true correlation that one could expect to find in an applicant population. Samples included in the meta-analysis consisted of employed individuals, meaning individuals who had been fired for counterproductive behavior of who had been screened on a correlated predictor (such as a Big Five inventory or an integrity test) would not be included in the samples, resulting in an attenuated correlation due to range restriction. Hiring managers may be able to improve their selection system by adding a measure of moral neutralization.

Of course, the utility of moral neutralization is not limited to the workplace. Across all of the studies, moral neutralization was a consistent and strong predictor of academic dishonesty. Cheating has always been a prevalent problem in academia (Whitley, 1998), and with the growing popularity of difficult-to-detect cheating methods, such as accessible large language models like ChatGPT, preventing academic dishonesty is more important today than it ever has been. Admissions committees could consider some kind of moral neutralization assessment when reviewing applicants. University admissions committees are increasingly moving away from traditional methods of assessing applicants and are looking for alternative, non-cognitive assessments (Hessen et al., 2022). A moral neutralization assessment may be an effective option decision-makers can use to fill this gap.

Limitations and Future Directions

Although the findings of the dissertation suggest that moral neutralization may be useful in applied settings, there are limitations to this work that should first be addressed. The most pressing of these limitations concerns the nature of moral neutralization measurement. All scales meant to be administered to adults measure exclusively moral disengagement or neutralization. Additionally, most of those measures are highly contextualized. Moral neutralization measures appear to predict counterproductive behavior even when there is a mismatch between the contextualized item content and outcome being predicted (Lee & Kuncel, 2020; Table 15). For example, Swan et al.'s (2017) driving moral disengagement measure correlates with academic dishonesty, but not as strongly as Haines et al's (1986) cheating neutralization scale or Moore et al.'s (2012) domain-neutral moral disengagement scale. The use of a composite formed out of several disparate assessments is a core limitation of this research and threatens the generalizability of the findings. For example, in Study 4, moral neutralization incrementally predicted academic dishonesty more effectively than infidelity intentions in both the truthful and fake responding conditions. However, that larger incremental prediction may be due to the inclusion of the contextualized Haines et al. measure. If that measure had been replaced with Lisman and Holman's (in press) measure of moral disengagement related to marital infidelity, the opposite pattern of effects may have been observed. Many of the measures that do exist have been subjected to very little validation research, especially the neutralization measures, which are often developed rationally for use in a single study. Scholars must work to address the issue of measurement by developing decontextualized measures that adequately sample the entire content domain of moral neutralization, rather than merely techniques of neutralization or mechanisms of moral disengagement.

Another limitation of this work is the emphasis on the overall moral neutralization construct rather than its specific facets. A goal of the dissertation was to merge the neutralization and moral disengagement literatures by gaining a deeper understanding of the broader moral neutralization construct than encompasses both. However, the lower-order structure of the construct is still unclear. Indeed, according to the Standards for Educational and Psychological Testing, collecting validity evidence based on internal structure, or "the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based" (p. 13), is an essential component of the eternal test validation process. Although the factor models investigated in Study 1 evaluated the internal structure of moral neutralization, no attempt was made to model the facet-level techniques of neutralization and mechanisms of moral neutralization that are the building blocks of their respective theories. In an earlier draft of Study 1, I prepared to fit models that included factors for each technique or mechanism, but there was insufficient data. For example, only two items measure the euphemistic labeling mechanism, so I did not have enough indicators to fit a model. In an attempt to address this issue, a group of undergraduate research assistants were given a sorting task with the goal of creating buckets of similar techniques and mechanisms, but agreement was near zero. Future research should address this limitation by collecting sufficent data on each facet and/or by ensuring that participants in the sorting task receive adequate training. It remains unclear which mechanisms and techniques are truly distinct from one another, and which are most important to moral neutralization measurement.

There is also disagreement as to whether technique/mechanism scores should be interpreted, or if all that matters is the overall score. Some have argued that the overall score should be used (Bandura et al., 1996; Moore et al., 2012), but others have found success focusing on specific mechanisms and techniques (e.g., Barsky, 2011; Betts et al., 2014). Like in other areas of individual differences measurement, it may be that relationships with other variables are maximized when the breadth of the predictor is matched to the breadth of the outcome (Ones & Viswesvaran, 1996). In other words, specific mechanisms and techniques might predict specific outcomes more effectively, but the broad moral neutralization construct might predict broad outcomes, like overall CWB, more effectively.

Another methodological limitation of this dissertation concerns the use of self-report measures. Nearly all variables were measured via simple self-report scales, so some correlations may be inflated by common method bias. Method variance is “variance that is attributable to the measurement method rather than to the construct of interest… At a more abstract level, method effects might be interpreted in terms of response biases such as halo effects, social desirability, acquiescence, leniency effects, or yea- and nay-saying” (Bagozzi & Yi, 1991, p. 426). When two variables are measured using the same method, their observed correlation may be inflated by their shared method variance (Campbell & Fiske, 1959; Podsakoff et al., 2003). Seeing as the moral neutralization, personality, and outcome measures used in this study were all self-report, common method bias may influence the effects. Indeed, the response biases that plague self-report assessments evidently affected that data used in these studies. For example, 10.5% of REP respondents failed the obvious attention check items. Although those respondents were removed from the analyses, attentiveness should not be viewed as a dichotomous variable. It is likely that some respondents paid just enough attention to pass the attention checks, but did not thoughtfully reply to all items.

Social desirability is another response bias that may have affected the observed results. Although the variable was not included in the primary analyses, the Balanced Inventory of Desirable Responding (BIDR) was administered to a subset of the respondents (N = 360). The BIDR is a 40-item measure that assesses two aspects of social desirability: impression management and self-deception. Social desirability was correlated with moral neutralization and the outcome variables included in the analyses (see Table 32). Impression management (r = -.51) was a stronger correlate of moral neutralization than self-deception (r = -.17). The same pattern was true for the academic dishonesty and CWB correlations, which suggests that respondents who want to make themselves look good will provide lower scores on each of the measures. A correlation with social desirability is not necessarily indicative of biased responding. Social desirability can be legitimately related to individual differences, including conscientiousness and neuroticism (Ones et al., 1996), traits that are associated with moral neutralization. Additionally, self-deception, a core component of social desirability, may be especially relevant to successfully invoking rationalizing attitudes (i.e., engaging in moral neutralization). However, the larger issue, that observed effects may be inflated due to shared method variance, remains.

Future research should address the flaw of shared method bias by measuring at least some of the variables with an alternative method. For example, instead of using a self-report measure of CWBs, researchers could gain access to organizational records to learn of past counterproductive behaviors. Alternatively, moral neutralization could be measured using another type of assessment, such as a situational judgment test (SJT). One of the original neutralization measures created by Ball (1966) used an SJT-like assessment with some success. Respondents were presented with vignettes describing a boy named Jack engaging in unethical behavior. For example, “Jack gets a club and goes with his friends to look for another group of boys. They find them in a park, and a fight starts. During the fight Jack hits another boy with the club, and almost kills him.” Adolescent respondents then rate the extent to which they agree with a series of justifications for Jack’s behavior, such as “People should not blame Jack this time if the other boy had once made him look like a coward.” Although the measure has been criticized by scholars (McCarthy & Stweart, 1998; Shields & Whitehall, 1994), it was used for decades with much success. A similar measure meant for adults could be developed to combat the issue of common method bias.

Although the findings of this research are promising, they must be interpreted with caution. Moral neutralization was strongly correlated with conceptually similar constructs. For example, the correlation with psychopathy was .55 (ρ = .64 after correcting for attenuation due to unreliability). Previous research has found similarly strong correlations. In a meta-analysis, Ogunfowora et al. (2021) found a reliability-correction mean correlation of .76 between psychopathy and moral disengagement (k = 17, N = 9,119; Table 13). Additionally, there are conceptual similarities between moral neutralization and these other constructs. For example, moral neutralization and psychopathy are both individual difference variables related to antisocial behavior. Additionally, some psychopathy scales include a rationalization dimension. The Psychopathic Personality Inventory (Lilienfeld & Andrews, 1996; PPI) includes a Blame Externalization subscale which measures the extent to which the respondent “Blames others and rationalizes their own transgressions” (Benning et al., 2003, p. 343). The PPI is a proprietary assessment, so I was not able to review the item content directly, but blame externalization is conceptually similar to moral neutralization.

Moral neutralization incrementally predicted outcomes after controlling for psychopathy and other predictors, a finding consistent with previous research (Ogunfowora et al., 2021). However, the short measure of psychopathy used in this study did not include a blame externalization subscale. It may be that the findings may have differed if blame externalization was included in the analyses. As is, it is possible that moral neutralization assessments measure a psychopathy-like construct that emphasizes rationalizations in a way that makes items easier to endorse without threatening one’s ego. A brief examination of the distributions of scores on both traits supports this possibility (Figure 15). Psychopathy produced a positively skewed distribution, with most respondents scoring near the minimum. On the other hand, moral neutralization was relatively normally distributed. Future research should further investigate the overlap between moral neutralization and psychopathy. It should be noted that even if moral neutralization and psychopathy are extremely similar constructs, psychopathy is a clinical construct, and so cannot be used to make selection decisions in many applied settings, including employment settings. Therefore, moral neutralization may have use in applied settings as a subclinical trait regardless of its similarity with psychopathy.

Conclusion

For over a century, scholars have theorized that rationalization protects people from experiencing psychic distress when they engage in immoral behavior (Freud, 1901). A core tenet of both neutralization theory (Sykes & Matza, 1957) and moral disengagement theory (1990) is that antisocial actors mitigate the guilt associated with their harmful behavior by invoking rationalizing attitudes. An abundance of research in both literatures demonstrates that people who are more likely to invoke such rationalizing attitudes are more likely to engage in deviant behavior. Indeed, the findings of this dissertation confirm that scores on measures of moral disengagement and neutralization are strongly correlated, exhibit similar relationships with other variables, and tap into a broad moral neutralization construct.

Moral neutralization is associated with counterproductive behavior across a range of settings, even when several other personality variables are controlled for, and in some cases, even when respondents are maximally distorting their responses to appear desirable. Moral neutralization even incrementally predicts counterproductive behavior after controlling for integrity, a similarly valid and conceptually similar trait. Although much more work needs to be done to achieve a thorough understanding of moral neutralization, particularly as it relates to moral neutralization measurement, these findings suggest that moral neutralization may be of use to practitioners who desire to minimize counterproductive behavior in their organizations.

Tables

Table 1

Table 1: Techniques of Neutralization and Mechanisms of Moral Disengagement

Techniques of Neutralization (Sykes and Matza, 1957)	Mechanisms of Moral Disengagement (Bandura et al., 1996)
Appeal to Higher Loyalties: [S]acrificing the demands of the larger society for the demands of the smaller social groups to which the delinquent belongs such as the sibling pair, the gang, or the friendship clique	**Moral Justification**: Detrimental conduct is made personally and socially acceptable by portraying it in the service of valued social or moral purposes.
-	**Euphemistic Language**: Through sanitized and convoluted verbiage, destructive conduct is made benign.
Denial of the Victim: The injury… is not really an injury; rather, it is a form of rightful retaliation or punishment.	**Advantageous Comparison**: [By comparing one's actions] with more reprehensible activities, injurious conduct can be rendered benign or appear to be of little consequence.
Condemnation of the Condemners: The delinquent shifts the focus of attention from his own deviant acts to the motives and behavior of those who disapprove of his violations. His condemners, he may claim, are hypocrites, deviants in disguise, or impelled by personal spite.	**Displacement of Responsibility**: [P]eople view their actions as springing from the social pressures or dictates of others rather than as something for which they are personally responsible.
Denial of Responsibility: In so far as the delinquent can define himself as lacking responsibility for his deviant actions, the disapproval of self or others is sharply reduced in effectiveness as a restraining influence.	**Diffusion of Responsibility**: Personal agency is obscured by diffusion of responsibility for detrimental conduct. Responsibility can be diffused by division of labor for a venture with different members performing subdivided aspects that seem harmless in themselves but harmful in its totality... Group decision making is another common practice, one that enables

	otherwise considerate people to behave inhumanely. When everyone is responsible, no one really feels responsible.
Denial of Injury: The delinquent frequently… feels that his behavior does not really cause any great harm despite the fact that it runs counter to law. Defense of Necessity (Minor 1981)	**Distorting Consequences**: When people pursue activities harmful to others for personal gain, or because of social inducements, they avoid facing the harm they cause, or they minimize it.
Metaphor of the Ledger (Klockars, 1974; Minor 1981): one who feels he has a sufficient supply of good to his credit can indulge in some evil without feeling guilty.	**Attribution of Blame**: The offender views themselves as a faultless victim who was forced to behave in a deviant manner by an outside, adversarial force.
Denial of Humanity (Alvarez, 1997): Victims are perceived as sub-human	**Dehumanization**: Dehumanization... divests people of human qualities or attributes bestial qualities to them.

Note. Definitions are taken from works cited otherwise noted. Metaphor of the Ledger was first described by Klockars, 1974, but the present definition comes from Minor, 1981. The Denial of Humanity definition is paraphrased from Alvarez, 1997.

Table 2

Table 2: CFA Model Fit Indices

	DF	AIC	BIC	RMSEA
Model 1 - One Factor	902	162334	162796	0.118
Model 2 - Two Factor	901	157594	158061	0.101
Model 3 - Five Factor	902	151782	152243	0.075
Model 4 - One Factor Bifactor	858	153225	153921	0.061
Model 5 - Two Factor Bifactor	857	152714	153415	0.057

Note. AIC was used to compare models. If BIC had been used instead, results would have been equivalent. Models are not nested so cannot be directly compared using RMSEA

Table 3

Table 3: Model 4 Standardized Factor Loadings

	g	**S_1**	**S_2**	**S_3**	**S_4**	**S_5**
Moore_1	0.535116	0.284254	0	0	0	0
Moore_2	0.529606	0.3768	0	0	0	0
Moore_3	0.671028	0.136847	0	0	0	0
Moore_4	0.433162	0.316198	0	0	0	0
Moore_5	0.497965	0.557466	0	0	0	0
Moore_6	0.356852	0.562153	0	0	0	0
Moore_7	0.420846	0.543499	0	0	0	0
Moore_8	0.291865	0.552525	0	0	0	0
Swan_1	0.462216	0	0.368184	0	0	0
Swan_2	0.331457	0	0.471148	0	0	0
Swan_3	0.41585	0	0.36408	0	0	0
Swan_4	0.47168	0	0.18122	0	0	0
Swan_5	0.479386	0	0.416241	0	0	0
Swan_6	0.498546	0	0.141813	0	0	0
Swan_7	0.28802	0	0.43615	0	0	0

Swan_8	0.307349	0	0.390572	0	0	0
Swan_9	0.285717	0	0.490313	0	0	0
Swan_10	0.395632	0	0.424419	0	0	0
Swan_11	0.421968	0	0.586566	0	0	0
Swan_12	0.388036	0	0.590049	0	0	0
Swan_13	0.441245	0	0.616921	0	0	0
Haines_1	0.40801	0	0	0.764408	0	0
Haines_2	0.517596	0	0	0.674732	0	0
Haines_3	0.393336	0	0	0.755971	0	0
Haines_4	0.516968	0	0	0.608525	0	0
Haines_5	0.484882	0	0	0.699183	0	0
Haines_6	0.514013	0	0	0.696844	0	0
Haines_7	0.567878	0	0	0.59317	0	0
Haines_8	0.361063	0	0	0.762747	0	0
Haines_9	0.436271	0	0	0.762232	0	0
Haines_10	0.442422	0	0	0.745762	0	0
Haines_11	0.438392	0	0	0.761461	0	0
Thurman_1	0.592779	0	0	0	0.243061	0

Thurman_2	0.648112	0	0	0	0.381719	0
Thurman_3	0.557927	0	0	0	0.37643	0
Thurman_4	0.671884	0	0	0	0.484838	0
Thurman_5	0.625494	0	0	0	0.415793	0
Thurman_6	0.549753	0	0	0	0.235377	0
Thurman_7	0.577603	0	0	0	0.437306	0
Agnew_1	0.535758	0	0	0	0	0.62956
Agnew_2	0.522603	0	0	0	0	0.564964
Agnew_3	0.579193	0	0	0	0	0.450146
Agnew_4	0.579481	0	0	0	0	0.628842
Agnew_5	0.502405	0	0	0	0	0.502904

Note. Model loadings for Model 4 in Study 1

Table 4

Table 4: Guttman Indeterminacy Index

	g1	g2	S1	S2	S3	S4	S5
Model 4 - One Factor Bifactor	.824	NA	.475	.618	.868	.422	.650
Model 5 - Two Factor Bifactor	.822	.887	.491	.450	.900	.588	.735

Note. S1 = Moore et al. (2012) specific factor; S2 = Swan et al. (2017) specific factor; S3 = Haines et al. (1986) specific factor. S4 = Agnew and Peters (1986) specific factor; S5 = Thurman (1984) specific factor.

Table 5

Table 5: Correlations between factor scores and criteria

Variable	g	S_1	S_2	S_3	S_4	S_5	CWB	CWBI	CWBO	AD	ITIS
g											
S_1	0.12										
S_2	0.11	0.21									
S_3	0.09	0.17	-0.11								
S_4	0.23	-0.18	-0.16	-0.12							
S_5	0.15	-0.15	-0.19	0.01	-0.1						
CWB	0.45	0.14	0.13	-0.02	0.11	0.05	(0.88)				
CWBI	0.29	0.2	0.19	0.01	0.09	0	0.79	(0.83)			
CWBO	0.43	0.09	0.09	-0.03	0.09	0.07	0.89	0.44	(0.84)		
AD	0.46	0.19	0.07	0.21	0.07	0.07	0.51	0.38	0.46	(0.91)	
ITIS	0.33	0.15	0.03	-0.02	0.17	0.08	0.33	0.28	0.3	0.36	(0.81)

Note. S1 = Moore et al. (2012) specific factor; S2 = Swan et al. (2017) specific factor; S3 = Haines et al. (1986) specific factor. S4 = Agnew and Peters (1986) specific factor; S5 = Thurman (1984) specific factor; CWB = Counterproductive work behaviors; CWBI = Individual CWBs; CWBO = Organizational CWBs; AD = Academic dishonesty; ITIS = Intentions Toward Infidelity Scale

Table 6

Table 6: Test of Measurement Invariance

	DF	AIC	BIC	χ^2	P-value
Configural	2574	152471	155255	7526.5	NA
Metric	2738	152757	154676	8140.2	< .001
Scalar	2814	152915	154433	8450.0	< .001

Note. Results are statistically significant, suggesting both metric and scalar measurement noninvariance.

Table 7

Table 7: Correlations Between Moral Neutralization and Workplace Outcomes

	k	***N***	$\bar{r}$	SD_r	$\bar{\rho}$	SD_{r_c}	SD_ρ	**95% CI**	**80% CV**	***I*²**
CWB	77	26,079	.42	.15	.45	.15	.14	[.42, .49]	[.27, .63]	90.55
Not Matched	54	16,057	.43	.16	.46	.16	.15	[.42, .50]	.26, .66]	90.65
Matched	19	8,248	.41	.13	.43	.13	.12	[.36, .49]	[.26, .59]	90.51
CWB-O	30	10,606	.38	.14	.40	.15	.14	[.35, .45]	[.22, .58]	89.70
CWB-I	19	6,535	.36	.10	.39	.11	.10	[.33, .44]	[.26, .51]	77.19
Specific CWB	16	4,522	.41	.18	.44	.18	.17	[.34, .53]	[.20, .67]	92.51
General CWB	61	21,557	.43	.14	.46	.14	.13	[.42, .49]	[.28, .63]	90.04
Unpublished	17	7,805	.45	.15	.49	.15	.14	[.41, .56]	[.29, .68]	93.74
Published	59	17,927	.41	.14	.44	.15	.14	[.40, .48]	[.26, .62]	89.14

	k	N	$\bar{r}$	SD_r	$\bar{\rho}$	SD_{r_c}	SD_ρ	**95% CI**	**80% CV**	I^2
UPB	11	2,043	.46	.14	.49	.16	.14	[.38, .59]	[.29, .68]	83.59
Turnover Intentions	5	1,617	.37	.13	.38	.13	.12	[.22, .54]	[.20, .56]	86.10
OCB	11	3,604	−.14	.18	−.15	.19	.18	[−.28, −.03]	[−.40, .10]	90.94
Job Satisfaction	4	1,186	−.21	.22	−.22	.23	.23	[−.59, .15]	[−.59, .15]	93.37

Note: k = number of studies contributing to meta-analysis; N = total sample size; $\bar{r}$ = mean observed correlation; SD_r = observed standard deviation of r; $\bar{\rho}$ = mean true-score correlation adjusted for measurement error; SD_{r_c} = observed standard deviation of corrected correlations ($\bar{\rho}$); SD_ρ = residual standard deviation of ρ; CI = confidence interval around $\bar{\rho}$; CV = credibility interval around $\bar{\rho}$; I^2 = estimated percent variance not accounted for by sampling error. CWB = counterproductive work behavior; CWB-O = organizational counterproductive work behavior; CWB-I = interpersonal counterproductive work behavior; Specific CWB = A specific counterproductive work behavior; UPB = Unethical Pro-Organizational Behavior; OCB = Organizational Citizenship Behavior. Correlations corrected individually.

[a]Negative values of I^2 are set equal to zero so that I^2 lies between .00% and 100% (Higgins et al., 2003)

Table 8

Table 8: Different Measures of Moral Disengagement Correlations with Counterproductive Work Behaviors

Measure	***k***	***N***	$\bar{r}$	SD_r	$\bar{\rho}$	SD_{r_c}	SD_ρ	**95% CI**	**80% CR**	I^2
All Levels	77	26,079	.42	.15	.45	.15	.14	[.42, .49]	[.27, .63]	90.55
Bandura	11	4,221	.41	.13	.45	.13	.12	[.37, .54]	[.29, .62]	89.70
Detert	8	1,942	.38	.05	.41	.06	.00	[.36, .46]	[.41, .41]	0.00[a]
Moore	29	7,950	.47	.17	.50	.17	.16	[.44, .57]	[.29, .72]	92.06
Neutralization	9	2,702	.49	.15	.51	.15	.14	[.39, .62]	[.30, .71]	91.26

Note: k = number of studies contributing to meta-analysis; N = total sample size; $\bar{r}$ = mean observed correlation; SD_r = observed standard deviation of r; $\bar{\rho}$ = mean true-score correlation adjusted for measurement error; SD_{r_c} = observed standard deviation of corrected correlations ($\bar{\rho}$); SD_ρ = residual standard deviation of ρ; CI = confidence interval around $\bar{\rho}$; CV = credibility interval around $\bar{\rho}$; I^2 = estimated percent variance not accounted for by sampling error; Bandura = Bandura et al. (1996); Detert = Detert et al. (2008); Moore = Moore et al. (2012). Correlations corrected individually. Not all Neutralization studies used the same measures.

[a] Negative values of I^2 are set equal to zero so that I^2 lies between .00% and 100% (Higgins et al., 2003)

Table 9

Table 9: Correlations with Counterproductive Work Behaviors Depending on the Theory

Predictor Theory	k	N	$\bar{r}$	SD_r	$\bar{\rho}$	SD_{r_c}	SD_ρ	95% CI	80% CR	I^2
All Levels	77	26,079	.42	.15	.45	.15	.14	[.42, .49]	[.27, .63]	90.55
Moral Disengagement	66	22,625	.42	.14	.45	.15	.14	[.41, .49]	[.27, .63]	90.48
Neutralization	13	4,547	.44	.14	.45	.14	.13	[.37, .54]	[.27, .63]	90.08

Note: k = number of studies contributing to meta-analysis; N = total sample size; $\bar{r}$ = mean observed correlation; SD_r = observed standard deviation of r; $\bar{\rho}$ = mean true-score correlation adjusted for measurement error; SD_{r_c} = observed standard deviation of corrected correlations ($\bar{\rho}$); SD_ρ = residual standard deviation of ρ; CI = confidence interval around $\bar{\rho}$; CV = credibility interval around $\bar{\rho}$; I^2 = estimated percent variance not accounted for by sampling error. Correlations corrected individually. Ks do not add to 50 because one study used a combination of both moral disengagement and neutralization.

Table 10

Correlations with Academic Dishonesty

Predictor Theory	***k***	***N***	$\bar{r}$	SD_r	$\bar{\rho}$	SD_{r_c}	SD_ρ	**95% CI**	**80% CR**	I^2
All Levels	38	11,407	.38	.15	.42	.17	.16	[.36, .48]	[.21, .63]	88.55
Moral Disengagement	16	4,913	.39	.14	.45	.17	.16	[.36, .54]	[.23, .66]	87.86
Neutralization	22	6,494	.36	.15	.40	.18	.17	[.32, .48]	[.18, .62]	89.26
Not Matched	10	3,188	.31	.10	.35	.13	.11	[.26, .44]	[.19, .50]	76.37
Matched	28	8,219	.40	.15	.45	.18	.17	[.38, .52]	[.22, .67]	89.89
Unpublished	11	3,159	.33	.17	.36	.20	.19	[.22, .49]	[.10, .62]	90.69
Published	27	8,248	.39	.13	.44	.16	.15	[.38, .51]	[.25, .64]	87.04

Note: k = number of studies contributing to meta-analysis; N = total sample size; $\bar{r}$ = mean observed correlation; SD_r = observed standard deviation of r; $\bar{\rho}$ = mean true-score correlation adjusted for measurement error; SD_{r_c} = observed standard deviation of corrected correlations ($\bar{\rho}$); SD_ρ = residual standard deviation of ρ; CI = confidence interval around $\bar{\rho}$; CV = credibility interval around $\bar{\rho}$; I^2 = estimated percent variance not accounted for by sampling error. Correlations corrected individually. Ks do not add to 50 because one study used a combination of both moral disengagement and neutralization.

Table 11

Table 11: Test of Excess Significance - Academic Dishonesty Analyses

Effects	**Expected Significant**	**Observed Significant**	***p*-value**
38	**34.7682**	**36**	**.3551**

Table 12

Table 12: Test of Excess Significance - CWB Analyses

Effects	Expected Significant	Observed Significant	*p*-value
77	**73.8260**	**76**	**.1627**

Table 13

Table 13: Correlations Between Moral Disengagement and Select Personality Traits from Ogunfowora (2021)

Trait	***k***	***N***	$\bar{r}$	SD_r	$\bar{\rho}$	SD_ρ	**95% CI**	**90% CR**
Conscientiousness	20	8,441	−.30	.22	−.38	.25	[−.49, −.27]	[−.78, .02]
Machiavellianism	21	10,525	.53	.13	.67	.15	[.61, .74]	[.43, .91]
Narcissism	17	8,770	.26	.07	.33	.11	[.28, .38]	[.18, .48]
Psychopathy	17	9,119	.62	.14	.76	.17	[.67, .84]	[.48, 1.03]

Note: k = number of studies contributing to meta-analysis; N = total sample size; $\bar{r}$ = mean observed correlation; SD_r = observed standard deviation of r; $\bar{\rho}$ = mean true-score correlation adjusted for measurement error; SD_{r_c} = observed standard deviation of corrected correlations ($\bar{\rho}$); SD_ρ = residual standard deviation of ρ; CI = confidence interval around $\bar{\rho}$; CV = credibility interval around $\bar{\rho}$.

Table 14

Table 14: Correlation Matrix Containing All Moral Neutralization Variables

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) MNeut_Comp	(0.97)						
(2) MNeut_Factor	0.94						
(3) MD_Moore	0.84	0.90	(0.83)				
(4) MD_Swan	0.72	0.64	0.58	(0.87)			
(5) N_Haines	0.69	0.63	0.51	0.28	(0.97)		
(6) N_Agnew	0.74	0.64	0.51	0.36	0.41	(0.88)	
(7) N_Thurman	0.82	0.78	0.62	0.52	0.43	0.55	(0.87)

Note. All correlations are are statistically significant at the .05 level; MNeut = Moral Neutralization; C = Conscientiousness; A = Agreeableness; N = Neuroticism; O = Openness to Experience; E = Extraversion; Mach = Machiavellianism; Narc = Narcissism; Psych = Psychopathy; ICAR = International Cognitive Ability Resource; AD = Academic Dishonesty; CWBI = Interpersonal Counterproductive Work Behaviors; CWBO = Organizational Counterproductive Work Behaviors; CWB = Counterproductive Work Behaviors. Alphas provided on the diagonal.

Table 15

Table 15: Correlations Between Moral Neutralization and Other Predictors and Outcomes

	MNeut_Comp	MNeut_Factor	MD_Moore	MD_Swan	N_Haines	N_Agnew	N_Thurman
Conscientious	-0.34	-0.34	-0.29	-0.17	-0.22	-0.26	-0.34
Agreeableness	-0.46	-0.48	-0.48	-0.37	-0.26	-0.29	-0.34
Neuroticism	0.28	0.3	0.27	0.16	0.18	0.20	0.26
Openness	-0.15	-0.18	-0.19	-0.13	-0.10	-0.05	-0.1
Extraversion	-0.12	-0.11	-0.12	-0.08	0.00	-0.11	-0.16
Orderliness	-0.26	-0.26	-0.21	-0.12	-0.18	-0.22	-0.26
Industriousness	-0.32	-0.33	-0.29	-0.17	-0.20	-0.23	-0.33
Politeness	-0.48	-0.52	-0.49	-0.35	-0.33	-0.29	-0.35
Compassion	-0.33	-0.33	-0.35	-0.3	-0.14	-0.22	-0.24
Volatility	0.29	0.32	0.29	0.20	0.19	0.20	0.25
Withdrawal	0.22	0.24	0.21	0.09	0.15	0.17	0.23
Open_Facet	-0.09	-0.11	-0.14	-0.14	-0.03	-0.02	-0.03
Intellect	-0.16	-0.19	-0.18	-0.08	-0.14	-0.06	-0.13
Assertiveness	-0.04	-0.03	-0.04	0.02	0.02	-0.06	-0.08
Enthusiasm	-0.18	-0.16	-0.17	-0.15	-0.02	-0.13	-0.20
Narcissism	0.22	0.26	0.25	0.24	0.17	0.08	0.12
Machiavellian	0.51	0.50	0.52	0.46	0.27	0.35	0.36
Psychopathy	0.55	0.56	0.52	0.44	0.35	0.37	0.43
Impulsivity	0.30	0.31	0.27	0.14	0.21	0.24	0.29
Cog_Abil	-0.05	-0.05	-0.04	-0.01	-0.13	-0.01	0.01

Age	-0.23	-0.26	-0.18	-0.05	-0.22	-0.23	-0.20
Gender	-0.17	-0.15	-0.17	-0.20	-0.07	-0.11	-0.09
Infidelity	0.33	0.34	0.32	0.21	0.15	0.26	0.33
Dishonesty	0.49	0.49	0.43	0.30	0.41	0.33	0.38
CWB	0.45	0.44	0.42	0.36	0.22	0.32	0.39
CWB_I	0.31	0.32	0.29	0.30	0.16	0.18	0.27
CWB_O	0.43	0.41	0.42	0.32	0.21	0.31	0.37

Note. CWB = Counterproductive Work Behaviors; CWBI = Interpersonal Counterproductive Work Behaviors; CWBO = Organizational Counterproductive Work Behaviors; MNeut = Moral Neutralization; MD = Moral Disengagement; N = Neutralization.

Table 16

Table 16: Correlations Between Outcomes and Predictor Variables

	MNeut_Comp	MNeut_Factor	Infidelity	ADishonesty	CWB	CWB_I	CWB_O
Conscientious	-0.34	-0.34	-0.13	-0.20	-0.28	-0.17	-0.26
Agreeableness	-0.46	-0.48	-0.23	-0.27	-0.35	-0.32	-0.30
Neuroticism	0.28	0.30	0.17	0.23	0.25	0.15	0.24
Openness	-0.15	-0.18	-0.07	-0.12	-0.12	-0.15	-0.08
Extraversion	-0.12	-0.11	0.00	-0.02	-0.12	-0.01	-0.15
Orderliness	-0.26	-0.26	-0.09	-0.12	-0.20	-0.14	-0.17
Industriousness	-0.32	-0.33	-0.14	-0.22	-0.29	-0.15	-0.28
Politeness	-0.48	-0.52	-0.29	-0.33	-0.35	-0.35	-0.28
Compassion	-0.33	-0.33	-0.14	-0.16	-0.26	-0.22	-0.23
Volatility	0.29	0.32	0.19	0.23	0.25	0.20	0.22
Withdrawal	0.22	0.24	0.14	0.19	0.21	0.09	0.22
Open_Facet	-0.09	-0.11	-0.05	-0.05	-0.08	-0.13	-0.04
Intellect	-0.16	-0.19	-0.06	-0.16	-0.13	-0.12	-0.09
Assertiveness	-0.04	-0.03	0.05	0.01	-0.05	0.05	-0.08
Enthusiasm	-0.18	-0.16	-0.05	-0.05	-0.17	-0.08	-0.18
Narcissism	0.22	0.26	0.19	0.15	0.06	0.12	0.02
Machiavellian	0.51	0.50	0.32	0.30	0.33	0.28	0.28
Psychopathy	0.55	0.56	0.43	0.39	0.43	0.39	0.36
Impulsivity	0.30	0.31	0.18	0.23	0.31	0.21	0.29
Cog_Abil	-0.05	-0.05	0.03	-0.14	-0.03	0.01	-0.06

Age	-0.23	-0.26	-0.07	-0.17	-0.01	0.00	0.01
Gender	-0.17	-0.15	-0.08	-0.09	-0.15	-0.11	-0.15
Infidelity	0.33	0.34	1	0.36	0.33	0.28	0.30
Dishonesty	0.49	0.49	0.36	1	0.51	0.38	0.46
CWB	0.45	0.44	0.33	0.51	1	0.79	0.89
CWB_I	0.31	0.32	0.28	0.38	0.79	1	0.44
CWB_O	0.43	0.41	0.30	0.46	0.89	0.44	1

Note. CWB = Counterproductive Work Behaviors; CWBI = Interpersonal Counterproductive Work Behaviors; CWBO = Organizational Counterproductive Work Behaviors; MNeut = Moral Neutralization; MD = Moral Disengagement; N = Neutralization.

Table 17

Table 17: Incremental Validity Predicting CWBs

	CWB		
	Null	Composite	Factor
ΔR^2	NA	0.041	0.036
Intercept	1.18**	1.19**	1.1**
		B	
Volatility	-0.02	-0.02	-0.02
Withdrawal	0.04	0.04	0.04
Compassion	-0.01	-0.01	-0.01
Politeness	-0.12*	-0.08	-0.07
Industriousness	-0.18**	-0.16**	-0.15**
Orderliness	0.05	0.07*	0.06*
Enthusiasm	0.04	0.04	0.03
Assertiveness	0.04	0.07*	0.06
Intellect	0.03	0.02	0.04
Open_Facet	-0.01	-0.01	0
Machiavellianism	0.1**	0.03	0.05
Narcissism	-0.1*	-0.12**	-0.12**
Psychopathy	0.28**	0.21**	0.19**
Impulsivity	0.21**	0.2**	0.22**
Age	0	0**	0**

Gender	-0.06	-0.06	-0.08*
MTurk	0.37**	0.33**	0.29**
Prolific	0.29**	0.22**	0.22**
MNeut	NA	0.27**	0.21**
N	1,384	1,383	1,351
R^2	0.279	0.32	0.315
R^2 95% CI	[.239, .319]	[.280, .360]	[.275, .355]
Adjusted R^2	269	.311	.305

Note. Table contains change in R squared (ΔR^2) and unstandardized regression weights. CWB = Counterproductive Workplace Behaviors. Composite = Unit weighted composite of the standardized moral neutralization scale scores. Factor = General factor extracted in Study 1. Standardized regression coefficients can be found in Appendix 14.

Table 18

Table 18: Incremental Validity Predicting CWB-Is

	CWBI		
	Null	Composite	Factor
ΔR^2	NA	0.015	0.006
Intercept	1.22**	1.23**	1.19**
		B	
Volatility	0.04	0.04	0.04
Withdrawal	-0.02	-0.02	-0.02
Compassion	0.01	0.01	0
Politeness	-0.18**	-0.15**	-0.14*
Industriousness	-0.06	-0.04	-0.04
Orderliness	0.03	0.04	0.03
Enthusiasm	0.04	0.04	0.04
Assertiveness	0.08	0.1*	0.09*
Intellect	-0.03	-0.04	-0.03
Open_Facet	-0.05	-0.05	-0.04
Machiavellianism	0.08*	0.04	0.05
Narcissism	-0.1*	-0.11**	-0.11**
Psychopathy	0.3**	0.25**	0.23**
Impulsivity	0.17**	0.16**	0.17**
Age	0.01**	0.01**	0.01**

Gender	-0.04	-0.04	-0.06
MTurk	0.32**	0.3**	0.27**
Prolific	-0.04	-0.08	-0.08
MNeut	NA	0.18**	0.13**
N	1,384	1,383	1,351
R^2	0.225	0.240	0.231
R^2 95% CI	[.187, .263]	[.201, .279]	[.192, .270]
Adjusted R^2	.215	.230	.220

Note. Table contains change in R squared (ΔR^2) and unstandardized regression weights. CWBI = Interpersonal Counterproductive Workplace Behaviors. Composite = Unit weighted composite of the standardized moral neutralization scale scores. Factor = General factor extracted in Study 1. Standardized regression coefficients can be found in Appendix 14.

Table 19

Table 19: Incremental Validity Predicting CWBOs

	Null	Composite	Factor
ΔR^2	NA	0.04	0.04
Intercept	1.18**	1.19**	1.08*
		B	
Volatility	-0.05	-0.05	-0.05
Withdrawal	0.08	0.08*	0.09*
Compassion	-0.01	-0.01	-0.01
Politeness	-0.13*	-0.08	-0.06
Industriousness	-0.24**	-0.21**	-0.21**
Orderliness	0.07	0.1*	0.09*
Enthusiasm	0.04	0.03	0.03
Assertiveness	0.03	0.06	0.06
Intellect	0.08	0.07	0.09*
Open_Facet	-0.01	-0.01	-0.01
Machiavellianism	0.1**	0.01	0.02
Narcissism	-0.1*	-0.12**	-0.12**
Psychopathy	0.27**	0.18**	0.16**
Impulsivity	0.24**	0.23**	0.25**
Age	0	0	0
Gender	-0.08	-0.08	-0.11*

MTurk	0.38**	0.33**	0.29**
Prolific	0.58**	0.5**	0.49**
MNeut	NA	0.31**	0.25**
N	1,384	1,383	1,351
R^2	0.257	0.297	0.297
R^2 95% CI	[.218, .296]	[.257, .337]	[.257, .337]
Adjusted R^2	.248	.288	.287

Note. Table contains change in R squared (ΔR^2) and unstandardized regression weights. CWBO = Organizational Counterproductive Workplace Behaviors. Composite = Unit weighted composite of the standardized moral neutralization scale scores. Factor = General factor extracted in Study 1. Standardized regression coefficients can be found in Appendix 14.

Table 20

Table 20: Incremental Validity Predicting Academic Dishonesty

	Academic Dishonesty		
	Null	Composite	Factor
ΔR^2	NA	0.08	0.073
Intercept	0.94**	0.95**	0.84**
		B	
Volatility	-0.03	-0.03	-0.02
Withdrawal	0.06*	0.06**	0.05*
Compassion	0.04	0.04	0.03
Politeness	-0.1*	-0.06*	-0.04
Industriousness	-0.03	-0.01	-0.01
Orderliness	0.03	0.05*	0.05*
Enthusiasm	0.06*	0.05**	0.05*
Assertiveness	-0.01	0.02	0.02
Intellect	-0.05	-0.05*	-0.04
Open_Facet	0	0	0.01
Machiavellianism	0.07**	0.01	0.02
Narcissism	-0.01	-0.03	-0.04
Psychopathy	0.17**	0.11**	0.1**
Impulsivity	0.06	0.05	0.06
Age	0*	0	0

Gender	-0.03	-0.03	-0.04
MTurk	0.08	0.05	0.03
Prolific	0.11*	0.05	0.06
MNeut	NA	0.23**	0.19**
N	1,384	1,383	1,351
R^2	0.209	0.289	0.282
R^2 95% CI	[.171, .247]	[.249, .329]	[.342, .422]
Adjusted R^2	.198	.279	.271

Note. Table contains change in R squared (ΔR^2) and unstandardized regression weights. Composite = Unit weighted composite of the standardized moral neutralization scale scores. Factor = General factor extracted in Study 1. Standardized regression coefficients can be found in Appendix 14.

Table 21

Table 21: Incremental Validity Predicting Infidelity Intentions

	Infidelity Intentions		
	Null	Composite	Factor
ΔR^2	NA	0.01	0.01
Intercept	0.25	0.21	0.19
		B	
Volatility	0.02	0.02	0
Withdrawal	0.09	0.08	0.08
Compassion	0.03	0.03	0.02
Politeness	-0.08	-0.05	-0.04
Industriousness	-0.05	-0.03	-0.03
Orderliness	0.04	0.06	0.05
Enthusiasm	0.13	0.12	0.14
Assertiveness	-0.08	-0.06	-0.07
Intellect	0.03	0.02	0.02
Open_Facet	-0.03	-0.03	0
Machiavellianism	0.14*	0.09	0.1
Narcissism	0.09	0.08	0.07
Psychopathy	0.58**	0.53**	0.5**
Impulsivity	-0.02	-0.02	-0.01
Age	0	0.01	0.01

Gender	0.05	0.04	0.03
MTurk	NA	NA	NA
Prolific	NA	NA	NA
MNeut	NA	0.17**	0.16**
N	650	649	632
R^2	0.203	0.213	0.213
R^2 95% CI	[.150, .256]	[.159, .267]	[.175, .251]
Adjusted R^2	.183	.192	.183

Note. Table contains change in R squared (ΔR^2) and unstandardized regression weights. CWB = Counterproductive Workplace Behaviors. Composite = Unit weighted composite of the standardized moral neutralization scale scores. Factor = General factor extracted in Study 1. Standardized regression coefficients can be found in Appendix 14.

Table 22

Table 22: Summary of Change in R^2

	R^2	Composite Added R^2 (ΔR^2)	Factor Added R^2 (ΔR^2)
CWB	.279	.320 (.041)	.315 (.036)
CWB-I	.225	.240 (.015)	.231 (.006)
CWB-O	.257	.297 (.040)	.297 (.040)
Academic Dishonesty	.209	.289 (.080)	.282 (.073)
Infidelity Intentions	.203	.213 (.010)	.213 (.010)

Note: All R^2 are statistically significant at the .01 level. Composite = Unit weighted composite of the standardized moral neutralization scale scores. Factor = General factor extracted in Study 1.; CWB = Counterproductive Work Behaviors; CWBI = Interpersonal Counterproductive Work Behaviors; CWBO = Organizational Counterproductive Work Behaviors

Table 23

Table 23: Correlations Between Model 4 Factors and Study 3 Variables

	MNeut	S_1	S_2	S_3	S_4	S_5
MD_Moore	0.9	0.36	0.09	0.01	-0.01	0.01
MD_Swan	0.64	0.11	0.82	-0.11	0.12	0
N_Haines	0.63	0	-0.11	0.83	0.01	0.08
N_Thurman	0.78	-0.12	0.09	-0.01	0.72	0.16
N_Agnew	0.64	-0.15	-0.01	0.05	0.2	0.82
Conscientiousness	-0.34	0.03	0.04	-0.03	-0.2	-0.1
Agreeableness	-0.48	-0.25	-0.16	0.02	0	-0.03
Neuroticism	0.3	0.01	-0.01	0.02	0.09	0.05
Openness	-0.18	-0.19	-0.05	0	0.05	0.07
Extraversion	-0.11	0.05	-0.03	0.08	-0.14	-0.07
Orderliness	-0.26	0.03	0.04	-0.05	-0.16	-0.11
Industriousness	-0.33	0.02	0.04	-0.01	-0.18	-0.08
Politeness	-0.52	-0.22	-0.09	-0.05	0.02	-0.01
Compassion	-0.33	-0.21	-0.17	0.07	-0.01	-0.04
Volatility	0.32	0.06	0.05	0.01	0.04	0.04
Withdrawal	0.24	-0.04	-0.05	0.02	0.12	0.06
Openness_Facet	-0.11	-0.19	-0.11	0.04	0.09	0.06
Intellect	-0.19	-0.13	0.03	-0.04	0.01	0.05
Assertiveness	-0.03	0.1	0.04	0.05	-0.11	-0.06
Enthusiasm	-0.16	0	-0.09	0.09	-0.14	-0.05

Narcissism	0.26	0.19	0.13	0.04	-0.11	-0.09
Machiavellianism	0.5	0.19	0.25	-0.03	0	0.08
Psychopathy	0.56	0.16	0.17	0.03	0.07	0.07
Impulsivity	0.31	0	-0.04	0.04	0.15	0.09
Cog_Abil	-0.05	-0.07	0.01	-0.12	0.1	0
Age	-0.26	0.03	0.13	-0.08	-0.03	-0.12
Gender	-0.15	-0.1	-0.15	0.02	0.03	-0.02
Infidelity	0.34	0.07	0.02	-0.05	0.17	0.09
Academic_Dishonesty	0.49	0.03	0.03	0.17	0.06	0.07
CWB	0.44	0.05	0.14	-0.05	0.14	0.09
CWB_I	0.32	0.13	0.16	-0.04	0.07	0
CWB_O	0.41	-0.01	0.11	-0.04	0.14	0.1

Note. S_1 through S_5 are the specific factors extracted in Study 1. S1 = Moore et al. (2012) specific factor; S2 = Swan et al. (2017) specific factor; S3 = Haines et al. (1986) specific factor. S4 = Agnew and Peters (1986) specific factor; S5 = Thurman (1984) specific factor; MNeut = Moral Neutralization general factor; MD = Moral Disengagement; N = Neutralization; CWB = Counterproductive Work Behaviors; CWBI = Interpersonal Counterproductive Work Behaviors; CWBO = Organizational Counterproductive Work Behaviors.

Table 24

Table 24: Correlation matrix of Study 4 Variables

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) MNeut.t	(.94)															
(2) MNeut.f	0.19	(.88)														
(3) Veris.t	-0.52	-0.11	(NA)													
(4) Veris.f	-0.10	-0.39	0.17	(NA)												
(5) Overt.t	-0.77	-0.07	0.54	0.08	(.93)											
(6) Overt.f	-0.11	-0.78	0.16	0.43	0.09	(.96)										
(7) Admit.t	0.52	0.10	-0.34	-0.07	-0.64	-0.10	(.86)									
(8) Admit.f	0.05	0.65	-0.10	-0.29	-0.04	-0.74	0.15	(.94)								
(9) CWB	0.43	0.02	-0.36	-0.04	-0.51	0.01	0.52	0.01	(.88)							
(10) AD	0.46	0.15	-0.28	-0.09	-0.45	-0.14	0.48	0.13	0.49	(.91)						
(11) Infidelity	0.33	0.15	-0.3	-0.09	-0.32	-0.08	0.36	0.06	0.33	0.36	(.81)					
(12) C	-0.39	-0.03	0.23	0.05	0.43	0.00	-0.26	-0.02	-0.32	-0.2	-0.13	(.91)				
(13) A	-0.44	-0.16	0.47	0.05	0.43	0.21	-0.3	-0.09	-0.36	-0.26	-0.23	0.39	(.89)			
(14) N	0.31	0.02	-0.17	-0.02	-0.37	0.00	0.23	-0.02	0.29	0.23	0.17	-0.53	-0.33	(.94)		
(15) O	-0.16	-0.10	0.06	-0.02	0.14	0.15	-0.08	-0.05	-0.13	-0.13	-0.07	0.31	0.44	-0.28	(.88)	
(16) E	-0.17	0.05	-0.11	-0.11	0.21	-0.12	-0.05	0.11	-0.14	-0.07	0.00	0.47	0.29	-0.54	0.43	(.93)

Note. Variables ending in .t were administered in the truthful responding block. Variables ending in .f were administered in the faked responding block. MNeut = Moral Neutralization; Veris = personality-based integrity; Admit = Admissions integrity; Overt = Overt integrity attitudes; CWB = counterproductive workplace behaviors; AD = academic dishonesty; infidelity = infidelity intentions; C = Conscientiousness; A = Agreeableness; N = Neuroticism; O = Openness; E = Extraversion

Table 25

Table 25: Cohen's d Values Comparing True Responses Versus Directed Faking

Measure	***d***	***95% CI***
MD_Moore	-.62	[-.70, -.54]
MD_Swan	-.83	[-.91, -.74]
N_Haines	-1.22	[-1.33, -1.10]
N_Agnew	-.72	[-.81, -.64]
N_Thurman	-.77	[-.86, -.69]
Mean Moral Neutralization	-.83	NA
Overt Integrity - Attitudes	1.24	[1.14, 1.35]
Overt Integrity - Admissions	-.56	[-.64, -.48]
Personality-Based Integrity	.31	[.20, .42]

Note. Mean Moral Neutralization contains the mean Cohen's d value for all moral neutralization measures. MD = Moral Disengagement; N = Neutralization.

Table 26

Table 26: Study 4 Regression Models Predicting CWBs

	Overt Attitudes			Overt Admissions			Veris			MNeut		
	Null	Truth	Fake	Null	Truth	Fake	Null	Truth	Fake	Null	Truth	Fake
ΔR^2	NA	.108**	.001	NA	.158**	.005**	NA	.037**	0	NA	.071**	.001
Intercept	2.84**	3.93**	2.72**	2.84**	1.66**	2.69**	3.09**	3.12**	3.08**	2.83**	2.2**	2.81**
	B (*b*)			*B*			*B*			*B*		
C	-.24**	-.12**	-.24**	-.24**	-.14**	-.23**	-.19**	-.15**	-.19**	-.24**	-.16**	-.24**
A	-.35**	-.15**	-.368*	-.35**	-.21**	-.34**	-.38**	-.22**	-.38**	-.35**	-.18**	-.34**
N	.17**	.11**	.17**	.17**	.11**	.17**	.15**	.11**	.15**	.17**	.13**	.17**
O	.03	-.03	.03	.03	.01	.04	.09	.06	.09	.03	.01**	.04
E	.12**	.1**	.13**	.12**	.06	.1**	.09	0	.09	.12**	.09**	.11**
MTurk	.45**	.44**	.45**	.45**	.38**	.43**	NA	NA	NA	.45**	.43**	.44**
Prolific	.37**	.31**	.36**	.37**	.25**	.36**	NA	NA	NA	.37**	.32**	.35**
MNeut/Integ	NA	-.52**	.04	NA	.53**	.09*	NA	-.01**	0	NA	.3**	.04
N	1,064	1,064	1,064	1,064	1,064	1,064	540	540	540	1,061	1,061	1,061
R^2	.191	.299	.192	.191	.349	.196	.171	.208	.171	.190	.261	.191
R^2 95% CI	[.149, .233]	[.253, .345]	[.150, .234]	[.149, .233]	[.303, .395]	[.154, .238]	[.114, 228]	[.148, .268]	[.303, .395]	[.148, 232]	[.217, .307]	[.148, .232]
Adjusted R^2	.186	.294	.186	.186	.344	.190	.163	.199	.161	.185	.256	.185

Note. Table contains change in R squared (ΔR^2) compared to the null model when truthful or faked responses to measures of integrity or moral neutralization are added to the model. Unstandardized regression weights for the Big Five and the source of the respondent are also provided. C = Conscientiousness; A = Agreeableness; N = Neuroticism; O = Openness to Experience; E = Extraversion. Standardized regression coefficients can be found in Appendix 14.

Table 27

Table 27: Study 4 Regression Models Predicting Academic Dishonesty

	Overt Attitudes			Overt Admissions			Veris			MNeut		
	Null	Truth	Fake	Null	Truth	Fake	Null	Truth	Fake	Null	Truth	Fake
ΔR^2	NA	.119**	.017**	NA	.157**	.025**	NA	.029**	.008*	NA	.143**	.028**
Intercept	1.74**	2.47**	2.08**	1.74**	.99**	1.52**	1.71**	1.72**	1.8**	1.73**	1.17**	1.65**
	B			*B*			*B*			*B*		
C	-.08**	0	-.08**	-.08**	-.02	-.07*	-.05	-.03	-.04	-.08**	0	-.07**
A	-.17**	-.04	-.14**	-.17**	-.08**	-.16**	-.15**	-.06	-.15**	-.17**	-.02	-.14**
N	.11**	.07**	.11**	.11**	.07**	.11**	.12**	.1**	.11**	.11**	.08**	.11**
O	-.03	-.08**	-.01	-.03	-.05	-.01	-.02	-.03	-.01	-.03	-.06*	-.02
E	.11**	.1**	.09**	.11**	.07**	.09**	.06	.01	.05	.11**	.08**	.09**
MTurk	.06	.05	.05	.06	.01	.03	NA	NA	NA	.06	.04	.04
Prolific	.06	.03	.07	.06	-.01	.05	NA	NA	NA	.06	.02	.02
MNeut/Integ	NA	-.35**	-.11**	NA	.33**	.14**	NA	-0**	-0**	NA	.27**	.13**
N	1,064	1,064	1,064	1,064	1,064	1,064	540	540	540	1,061	1,061	1,061
R^2	.109	.228	.126	.109	.266	.134	.104	.133	.112	.108	.251	.136
R^2 95% CI	[.074, .144]	[.184, .272]	[.089, .163]	[.074, .144]	[.221, .311]	[.096, .172]	[.056, .152]	[.081, .185]	[.063, .161]	[.073, .143]	[.206, .296]	[.098, .174]
Adjusted R^2	.103	.222	.120	.103	.261	.127	.100	.123	.102	.102	.245	.130

Note. Table contains change in R squared (ΔR^2) compared to the null model when truthful or faked responses to measures of integrity or moral neutralization are added to the model. Unstandardized regression weights for the Big Five and the source of the respondent are also provided. C = Conscientiousness; A = Agreeableness; N = Neuroticism; O = Openness to Experience; E = Extraversion. Standardized regression coefficients can be found in Appendix 14.

Table 28

Table 28: Study 4 Regression Models Predicting Infidelity Intentions

	Overt Attitudes			Overt Admissions			Veris			MNeut		
	Null	Truth	Fake	Null	Truth	Fake	Null	Truth	Fake	Null	Truth	Fake
ΔR^2	NA	.044**	.001	NA	.076**	.001	NA	.035**	.005	NA	.048**	.013**
Intercept	2.73**	3.71**	2.91**	2.73**	1.53**	2.63**	2.52**	2.59**	2.7**	2.73**	2.05**	2.61**
	B			*B*			*B*			*B*		
C	-.06	.04	-.06	-.06	.02	-.05	-.03	.02	-.02	-.06	.03	-.06
A	-.42**	-.25**	-.41**	-.42**	-.28**	-.42**	-.37**	-.16	-.37**	-.42**	-.25**	-.39**
N	.24**	.17**	.24**	.24**	.18**	.24**	.26**	.21**	.26**	.24**	.19**	.24**
O	.05	-.01	.06	.05	.02	.06	.05	.01	.05	.05	.02	.07
E	.26**	.24**	.25**	.26**	.19**	.25**	.22**	.1	.2*	.26**	.22**	.23**
MNeut/Integ	NA	-.45**	-.05	NA	.49**	.06	NA	-.01**	-0**	NA	.33**	.2**
N	662	662	662	662	662	662	491	491	491	662	662	662
R^2	.09	.134	.091	.09	.166	.091	.084	.119	.089	.09	.138	.103
R^2 95% CI	[.049, .131]	[.086, .182]	[.050, .132]	[.049, .131]	[.115, .217]	[.050, .132]	[.038, .130]	[.066, .172]	[.042, .136]	[.049, .131]	[.090, .186]	[.060, .147]
Adjusted R^2	.083	.126	.082	.083	.158	.083	.074	.108	.078	.083	.130	.094

Note. Table contains change in R squared (ΔR^2) compared to the null model when truthful or faked responses to measures of integrity or moral neutralization are added to the model. Unstandardized regression weights for the Big Five and the source of the respondent are also provided. C = Conscientiousness; A = Agreeableness; N = Neuroticism; O = Openness to Experience; E = Extraversion. Standardized regression coefficients can be found in Appendix 14.

Table 29

Table 29: Incremental Validity of Moral Neutralization Over Integrity

Overt Attitudes						
	CWB		Academic Dishonesty		Infidelity Intentions	
	Null	Add MNeut	Null	Add MNeut	Null	Add MNeut
ΔR^2	NA	.011**	NA	.049**	NA	.016**
Intercept	4.17**	3.60**	2.68**	1.92**	4.38**	3.45**
	B		*B*		*B*	
Integrity	-.67**	-.50**	-.38**	-.17**	-.56**	-.29**
MNeut	NA	.16**	NA	.21**	NA	.26**
N	1,074	1,074	1,074	1,074	666	666
R^2	.245	.256	.201	.250	.101	.117
R^2 95% CI	[.200, .290]	[.211, .301]	[.158, .244]	[.205, .295]	[.058, .144]	[.071, .163]
Adjusted R^2	.244	.254	.201	.258	.099	.114
Admissions						
	CWB		Academic Dishonesty		Infidelity Intentions	
	Null	Add MNeut	Null	Add MNeut	Null	Add MNeut
ΔR^2	NA	.04**	NA	.074**	NA	.027**
Intercept	.82**	1.06**	.75**	.95**	1.5**	1.75**
	B		*B*		*B*	

Integrity	.65**	.50**	.38**	.25**	.58**	.43**
MNeut	NA	.23**	NA	.19**	NA	.25
N	1,074	1,074	1,074	1,074	666	666
R^2	.278	.318	.236	.310	.129	.156
R^2 95% CI	[.233, .323]	[.272, .364]	[.192, .280]	[.264, .356]	[.082, .176]	[.106, .206]
Adjusted R^2	.278	.317	.235	.309	.128	.153
Personality-Based						
	CWB		Academic Dishonesty		Infidelity Intentions	
	Null	Add MNeut	Null	Add MNeut	Null	Add MNeut
ΔR^2	NA	.088**	NA	.136**	NA	.04**
Intercept	2.37**	2.11**	1.58	1.38**	2.99**	2.75**
	B		*B*		*B*	
Integrity	-.01**	-.01**	-.01	-0**	-.01**	-.01**
MNeut	NA	.32**	NA	.25**	NA	.29**
N	609	609	609	609	557	557
R^2	.127	.215	.079	.215	.088	.128
R^2 95% CI	[.078, .177]	[.158, .272]	[.038, .120]	[.158, .272]	[.043, .133]	[.077, .179]
Adjusted R^2	.125	.212	.077	.212	.087	.125
Overt Attitudes and Admissions						
	CWB		Academic Dishonesty		Infidelity Intentions	
	Null	Add MNeut	Null	Add MNeut	Null	Add MNeut

ΔR^2	NA	.008**	NA	.041**	NA	.014**
Intercept	2.41**	1.97**	1.63	.98**	2.63**	1.79**
	B		*B*		*B*	
Admission	.44**	.43**	.26	.25**	.43**	.42**
Attitudes	-.36**	-.23**	-.20	-.01	-.26**	-.01
MNeut	NA	.13**	NA	.19**	NA	.24**
N	1,074	1,074	1,074	1,074	666	666
R^2	.321	.329	.269	.310	.142	.156
R^2 95% CI	[.275, .367]	[.283, .375]	[.224, .314]	[.264, .356]	[.093, .191]	[.106, .206]
Adjusted R^2	.320	.327	.268	.308	.139	.152

Note. Table contains change in R squared (ΔR^2) when moral neutralization is added to regression models that already includes integrity as a predictor. CWB = Counterproductive Workplace Behaviors

Table 30

Table 30: Descriptives of Moral Neutralization and Integrity Used in Study 4

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
MNeut	1,072	0.018	0.8	-1.5	-0.59	0.56	2.5
F.Mneut	1,072	-0.0011	0.69	-0.95	-0.45	0.31	2.8
Moore	1,072	2.5	1.1	1	1.6	3.2	6.2
F.Moore	1,072	1.8	1.2	1	1	2.1	7
Swan	1,072	2.7	0.75	1	2.2	3.2	4.8
F.Swan	1,072	2	0.89	1	1.2	2.5	5
Haines	1,072	1.9	0.99	1	1	2.5	5
F.Haines	1,072	3.6	1.6	1	2	5	5
Thurman	1,072	2	0.66	1	1.6	2.6	4
F.Thurman	1,072	1.5	0.67	1	1	2	4
Agnew	1,072	1.7	0.6	1	1	2.2	3
F.Agnew	1,072	1.3	0.5	1	1	1.4	3
Attitudes	1,076	3.4	0.57	1.7	3.1	3.9	4.9
F.Attitudes	1,076	4.2	0.65	1.2	3.9	4.7	4.9
Admissions	1,076	1.6	0.62	1	1.2	1.8	4.7
F.Admissions	1,076	1.3	0.59	1	1	1.2	5
Personality	544	41	22	0	25	56	96
F.Personality	544	48	26	0	29	70	97

Note. Variables beginning with F. were administered during the faking condition. All other variables were measured in the truthful responding condition.MNeut = Moral Neutralization; Attitudes = Overt Integrity Attitudes; Admissions = Overt Integrity Admissions; Personality = Personality-Based Integrity.

Table 31

Table 31: Correlations Between Cognitive Ability and Score Improvement in the Faking Condition

	ICAR
Moore_diff	0.13
Swan_diff	0.16
Haines_diff	0.1
Thurman_diff	0.16
Agnew_diff	0.16
Attitudes_diff	0.19
Admissions_diff	0.15
Personality_diff	0.13

Note. Pairwise Ns range from 684 to 686, with the exception of the correlation with the personality-based integrity test (N = 258). A positive correlation indicates association with an improved score (more integrity or less moral neutralization). All correlations were statistically significant at the .05 level. Attitudes = Overt Integrity Attitudes; Admissions = Overt Integrity Admissions; Personality = Personality-Based Integrity.

Table 32

Table 32: Correlations Between Social Desirability, Moral Neutralization, and Outcomes

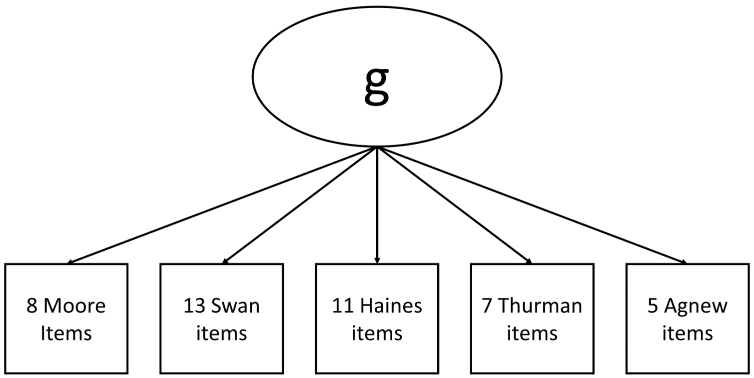
	BIDR	Self_Deception	Impression_Management
Moral_Neutralization	-0.43	-0.17	-0.51
Academic_Dishonesty	-0.3	-0.16	-0.33
CWB	-0.43	-0.23	-0.45

Note. N = 360 because only a subset of the sample was administered the BIDR. No participants were administered the BIDR and the ITIS, so correlations between social desirability and infidelity intentions could not be computed. BIDR = Balanced Inventory of Desirable Responding. CWB = Counterproductive Workplace Behavior.

Figures

Figure 1

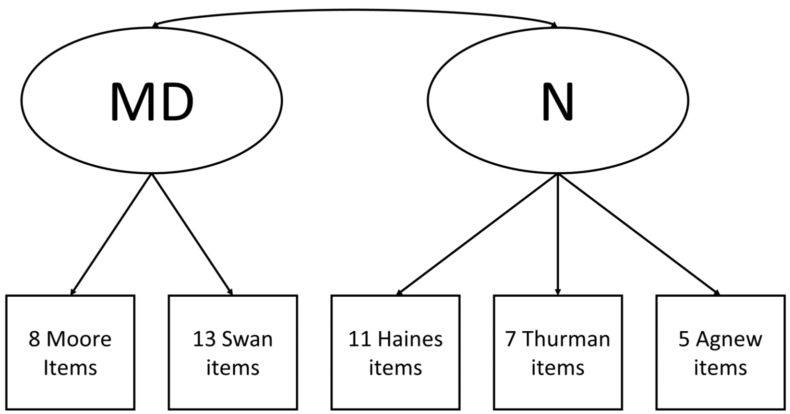
Figure 1: Model 1



Note. The 1-factor model. Indicators have been grouped to increase readability. The model will be run with all 44 individual items as indicators.

Figure 2

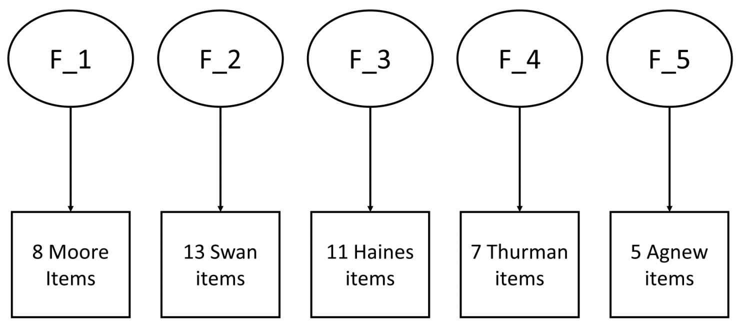
Figure 2: Model 2



Note. The 2-factor model. Indicators have been grouped to increase readability. The model will be run with all 44 individual items as indicators.

Figure 3

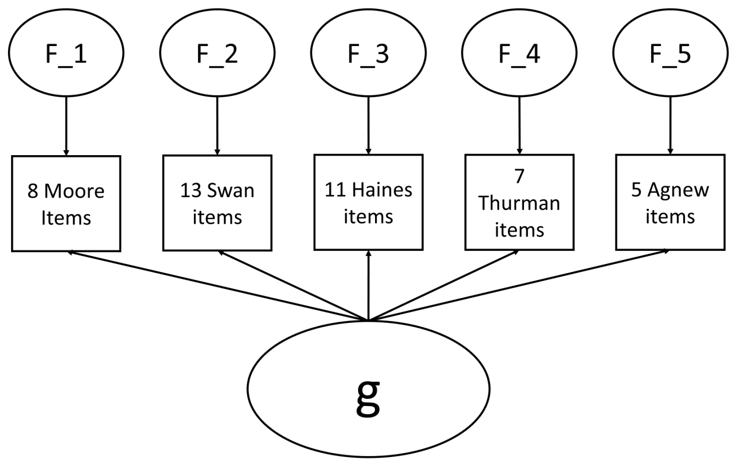
Figure 3: Model 3



Note. The 5 uncorrelated factors model. Indicators have been grouped to increase readability. The model will be run with all 44 individual items as indicators.

Figure 4

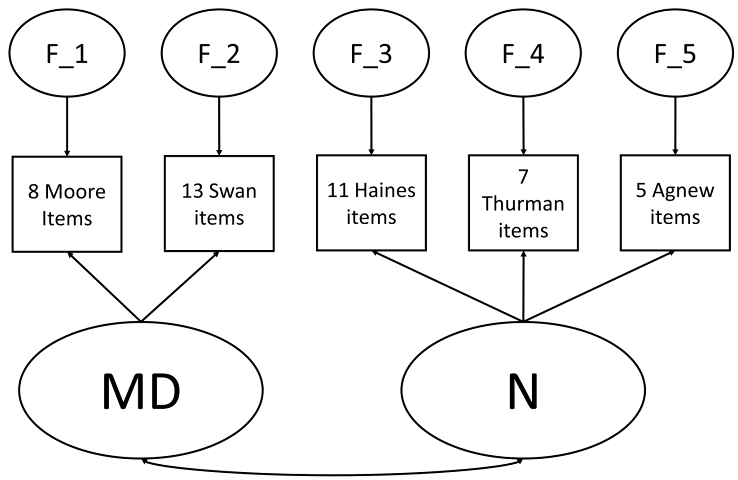
Figure 4: Model 4



Note. The bifactor model. Indicators have been grouped to increase readability. The model will be run with all 44 individual items as indicators.

Figure 5

Figure 5: Model 5



Note. The two general factors and five specific factors model. Indicators have been grouped to increase readability. The model will be run with all 44 individual items as indicators.

Figure 6

Figure 6: Flowchart describing the workplace outcome data collection process

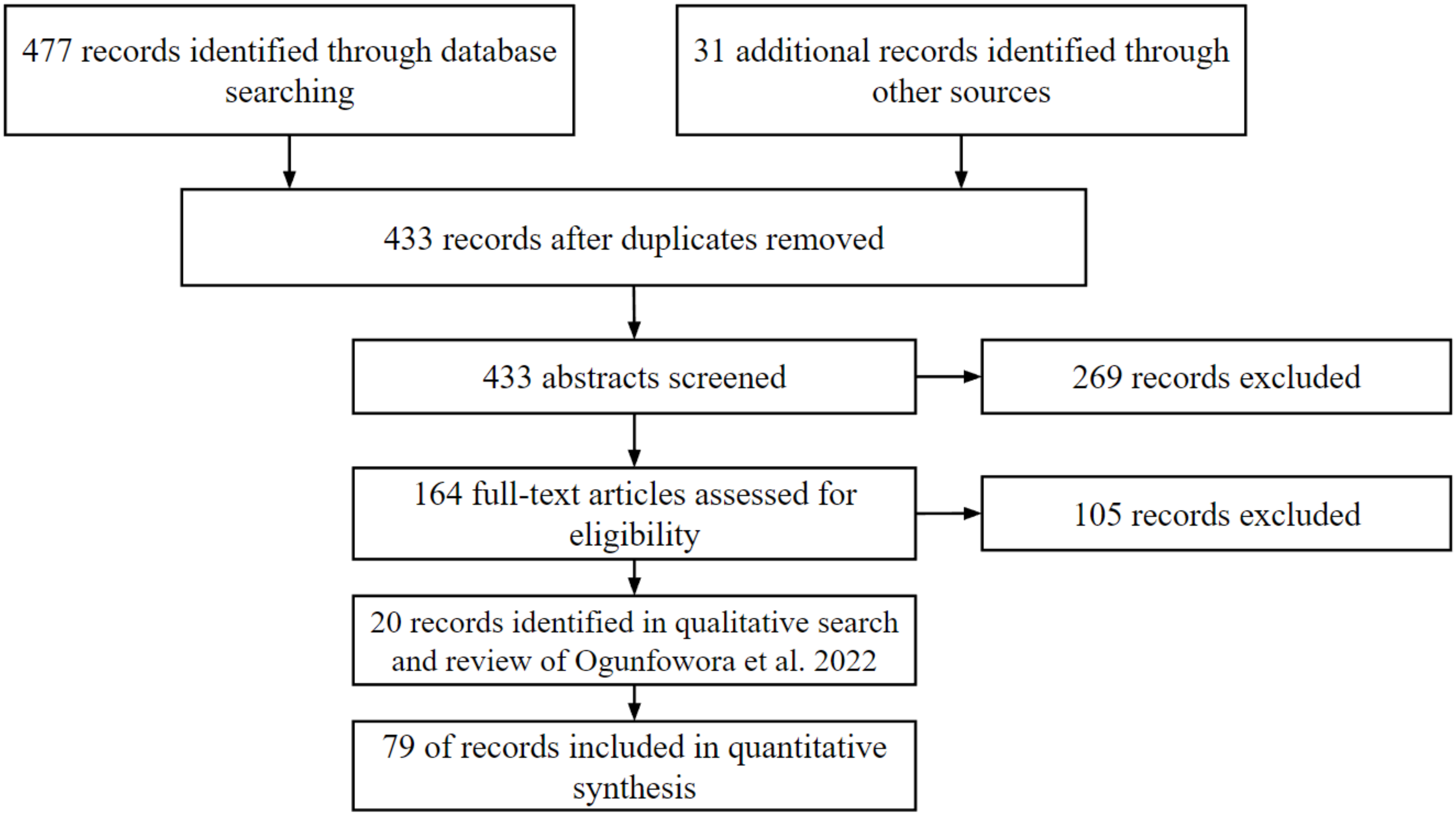


Figure 7

Figure 7: Flowchart describing the academic dishonesty data collection

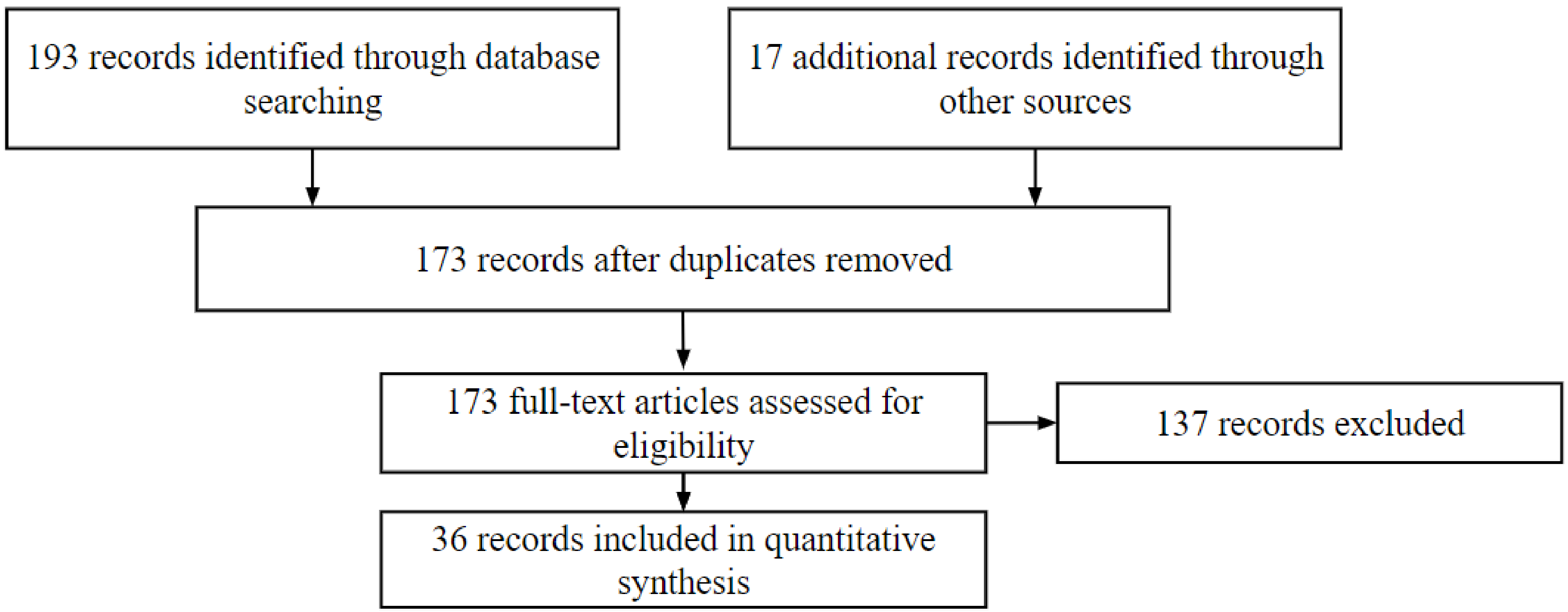
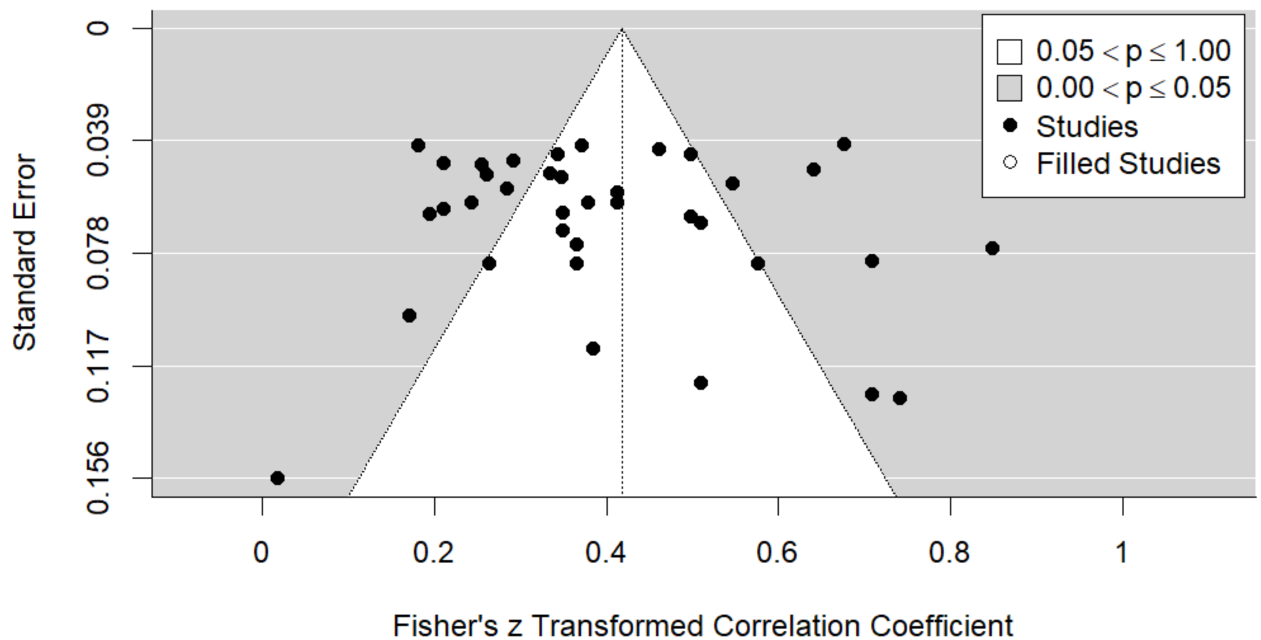


Figure 8

Figure 8: Academic Dishonesty Trim and Fill



Note. No asymmetry was detected, so no effects were imputed.

Figure 9

Figure 9: Academic Dishonesty PET-PEESE

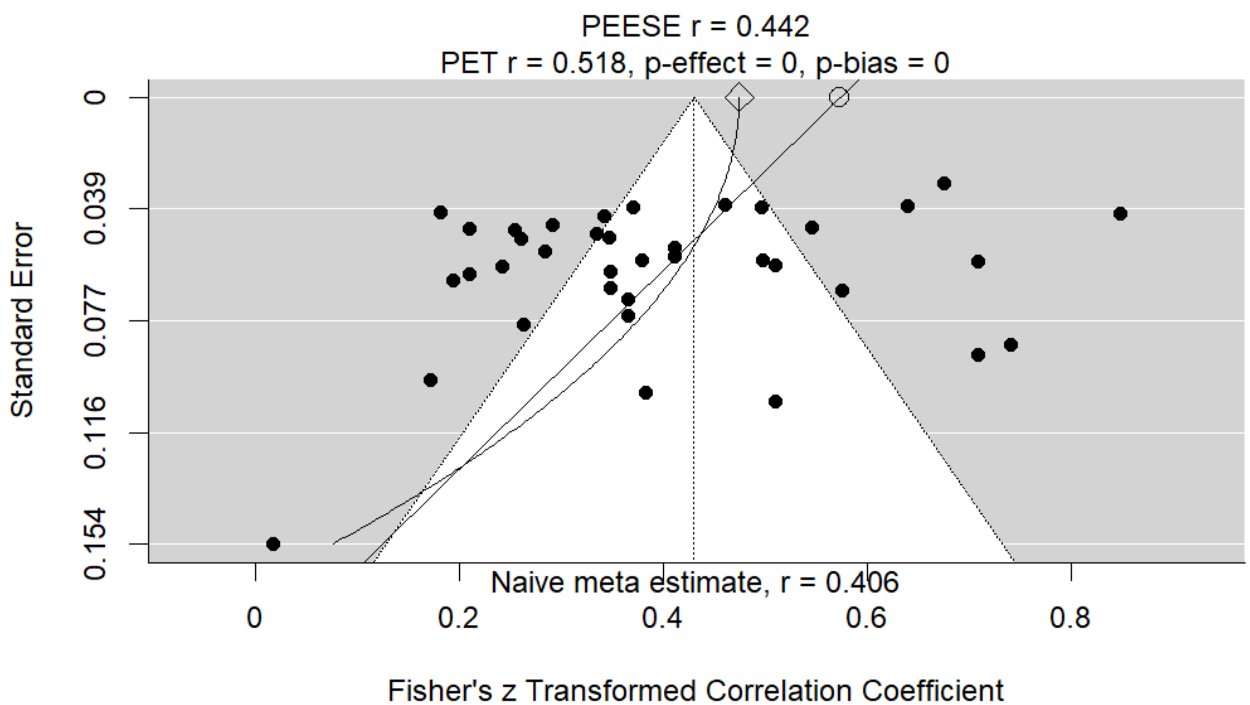


Figure 10

Figure 10: Academic Dishonesty Cumulative Meta-analysis

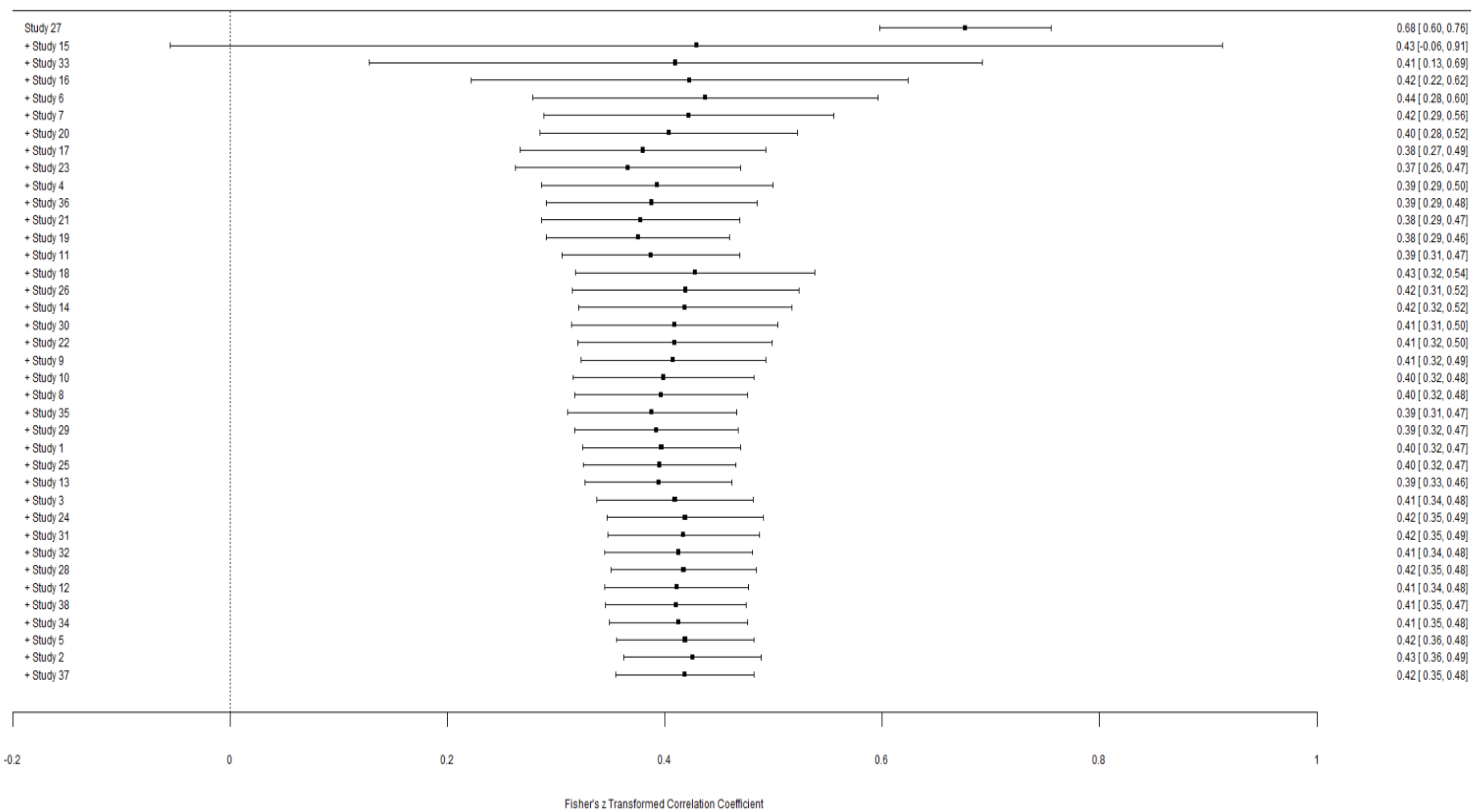
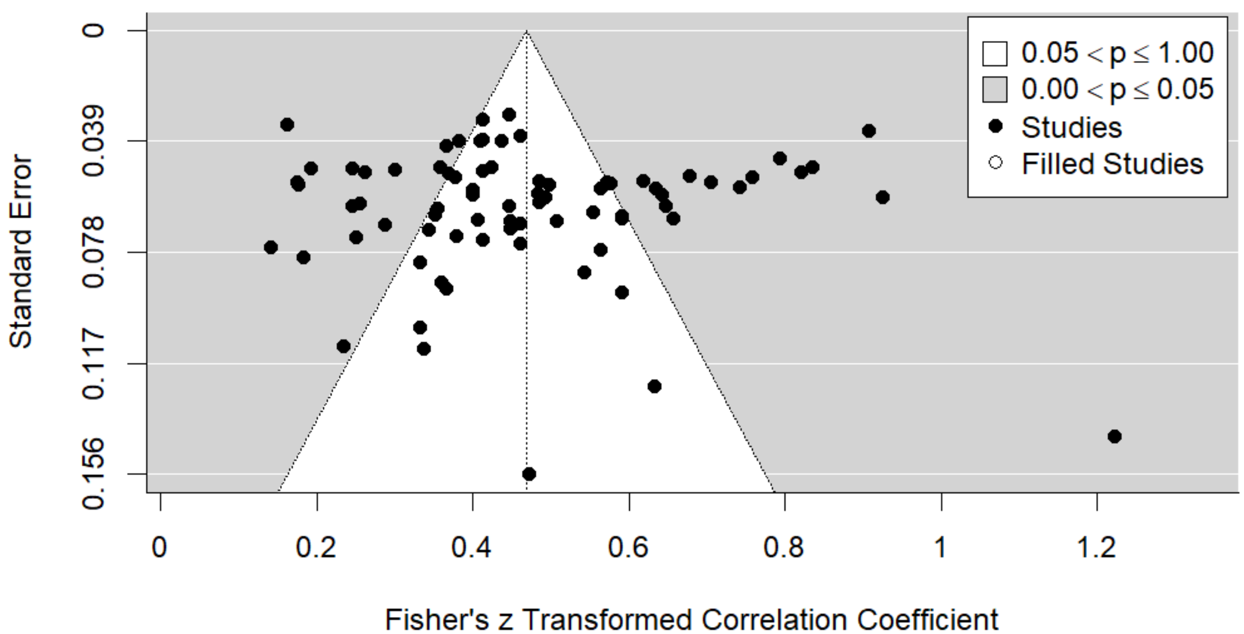


Figure 11

Figure 11: CWB Trim-and-Fill



Note. No asymmetry was detected, so no effects were imputed

Figure 12

Figure 12: CWB PET-PEESE

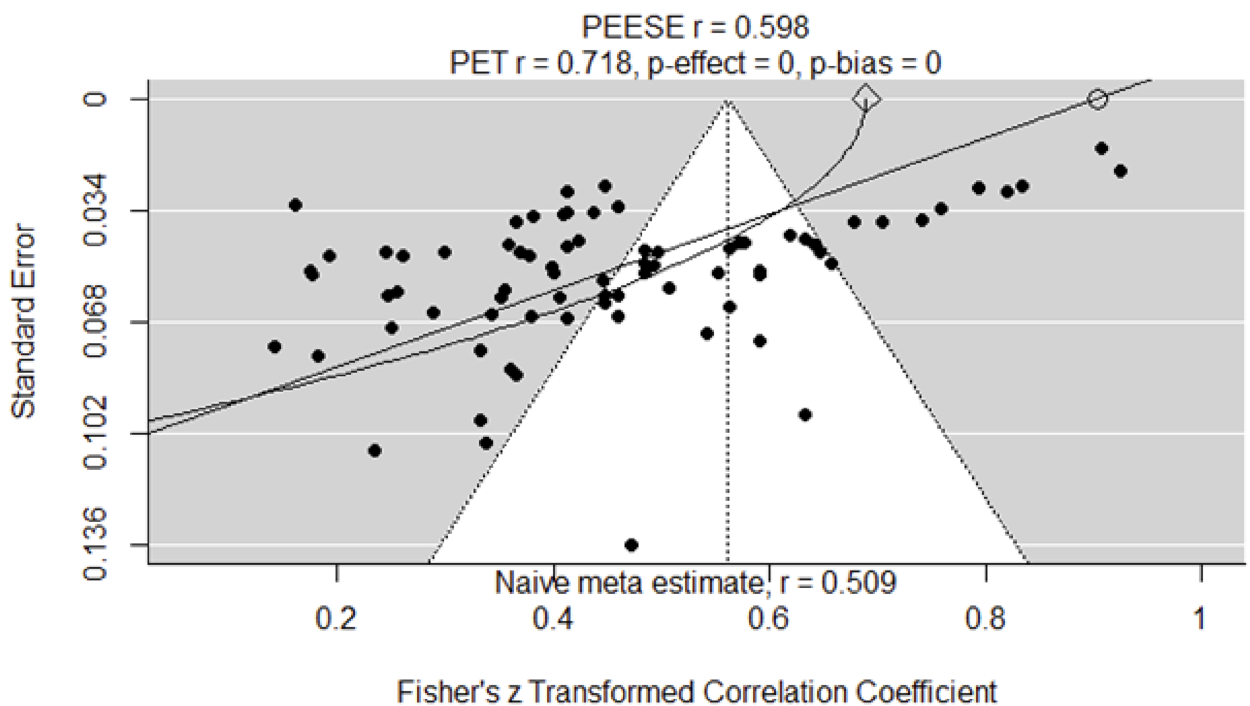
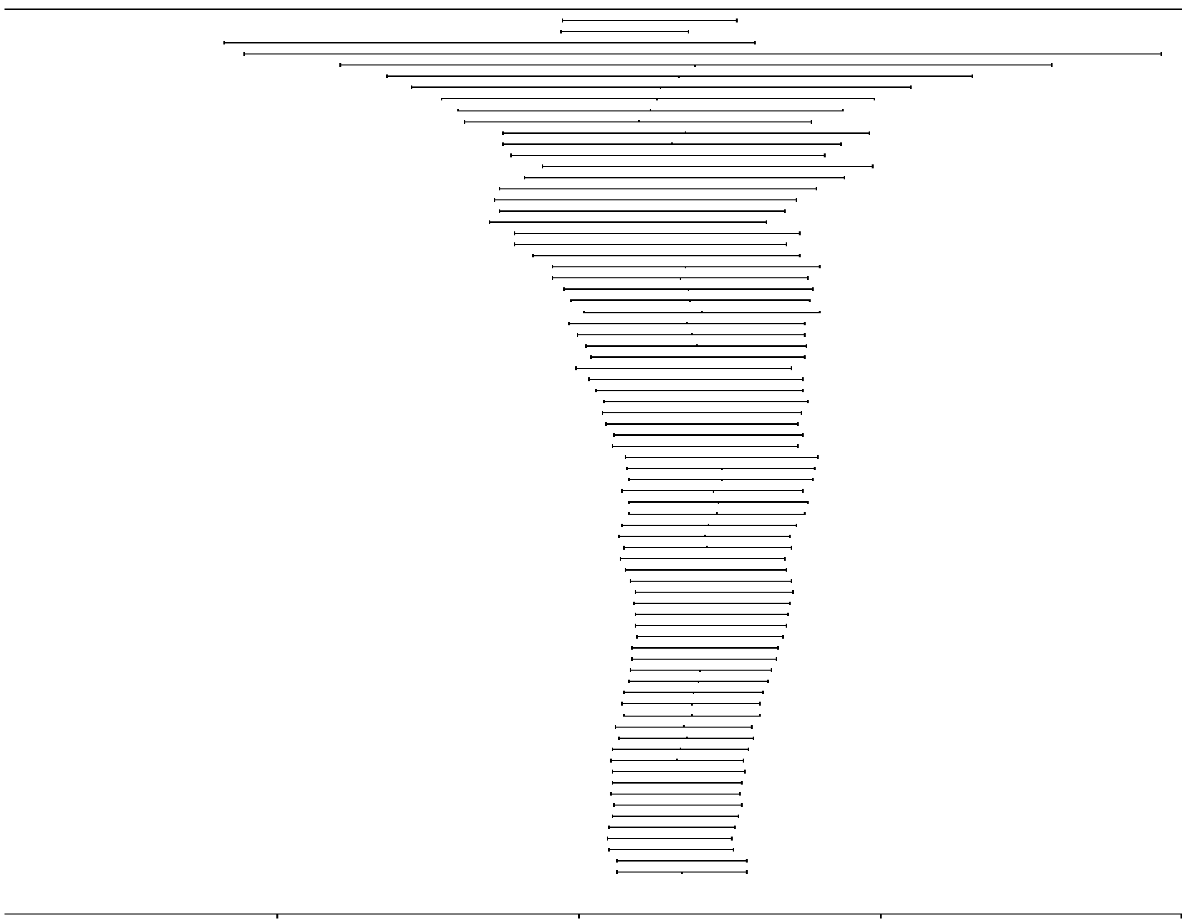


Figure 13

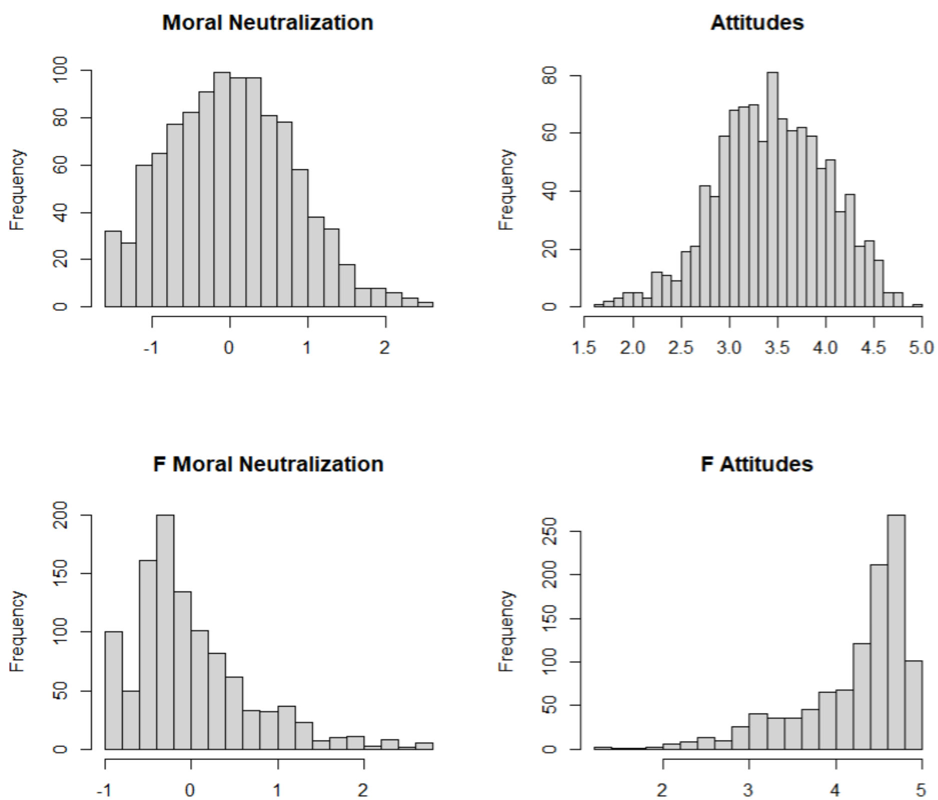
Figure 13: CWB Cumulative Meta-Analysis



Note. Effect sizes not shown due to limited space. Effect size = Fisher's z-transformed correlation coefficients. 0 Not included in any intervals

Figure 14

Figure 14: Histograms of Integrity and Moral Neutralization Scores



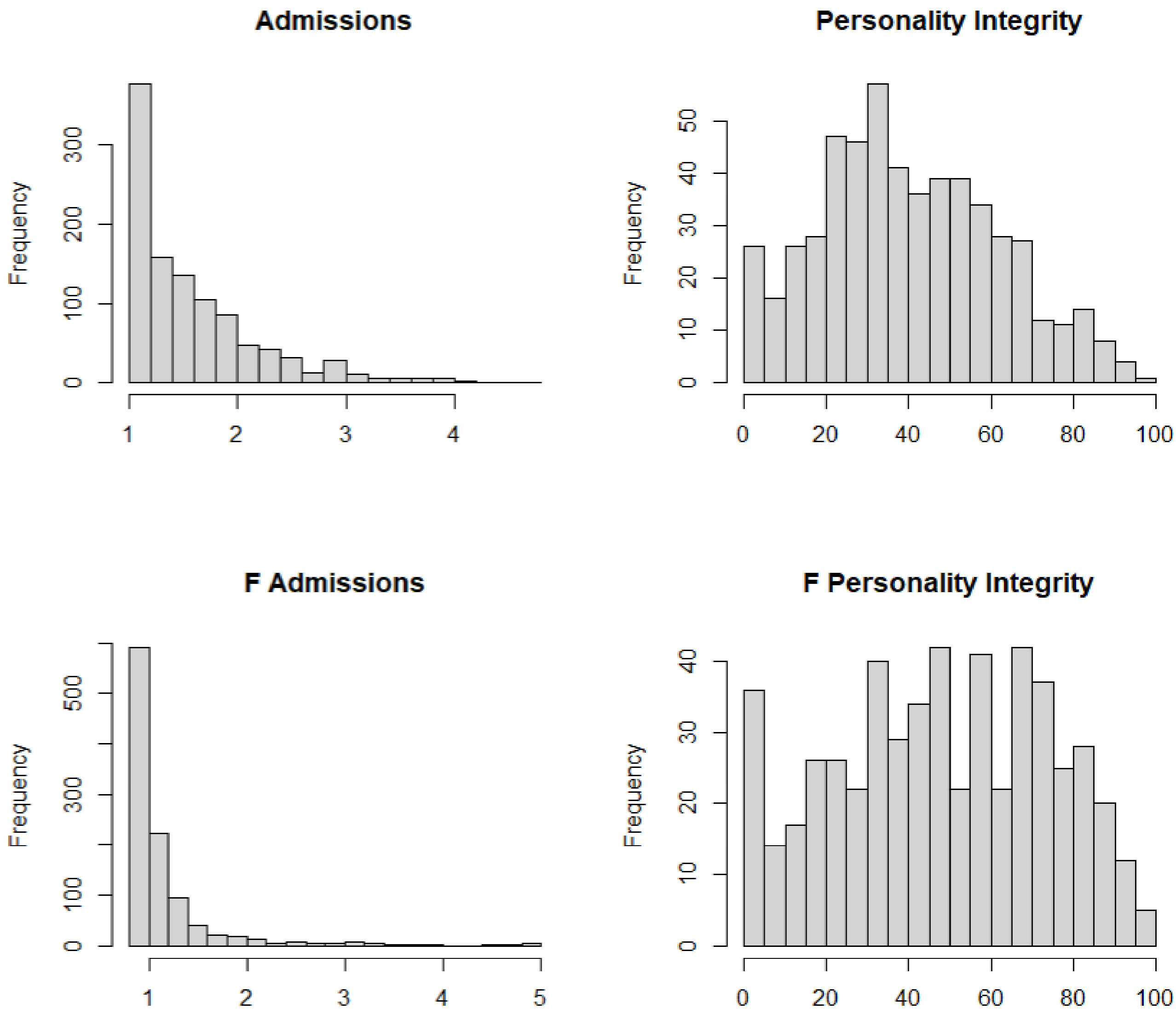
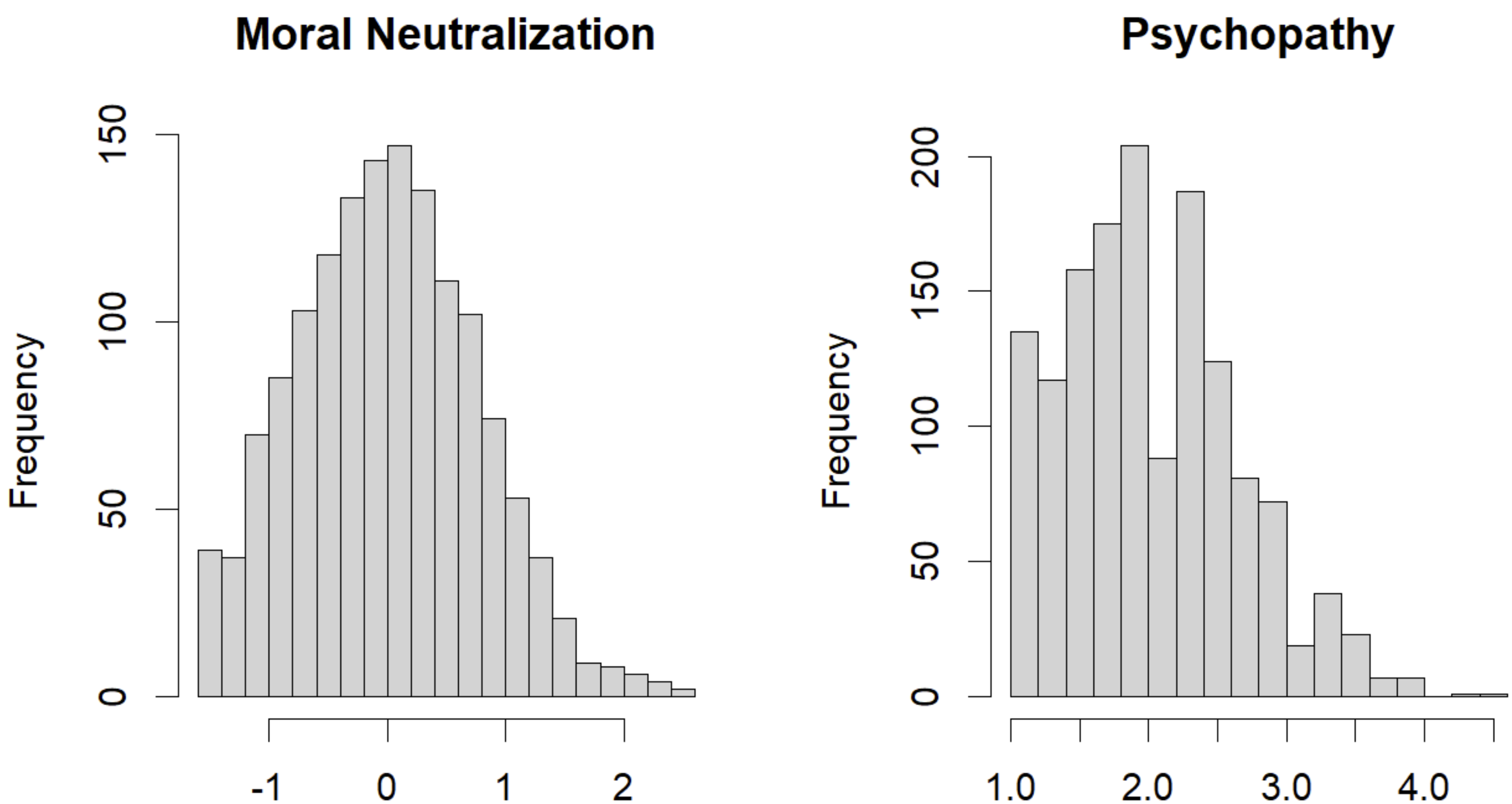

Admissions
Frequency
0
100
200
300
1
2
3
4
Personality Integrity
Frequency
0
10
20
30
40
50
0
20
40
60
80
100
F Admissions
Frequency
0
100
300
500
1
2
3
4
5
F Personality Integrity
Frequency
0
10
20
30
40
0
20
40
60
80
100

Figure 15

Figure 15: Histogram of Moral Neutralization and Psychopathy Scales



Bibliography

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *The standards for educational and psychological testing* (2nd ed.). Washington, DC: American Educational Research Association.

Anderson, D., & Burnham, K. (2006). *Aic myths and misunderstandings*. https://sites.warnercnr.colostate.edu/anderson/wp-content/uploads/sites/26/2016/11/AIC-Myths-and-Misunderstandings.pdf.

Agnew, R. & Peters, A. A. R. (1986). The techniques of neutralization: An analysis of predisposing and situational factors. *Criminal Justice and Behavior,* 81-97.

Alvarez, A. (1997) Adjusting to genocide: The techniques of neutralization and the holocaust. *Social Science History, 21*(2), 139-178.

*Andrews, J. S. (2018). *A Tale of Two Constructs: Defining Integrity and Corruption, Clarifying Their Factor Structures and Examining Their Association with Organizational Citizenship and Counterproductive Work Behaviors* [Unpublished master's thesis]. North Carolina State University.

Bagozzi, R. P., & Yi, Y. (1991). Multitrait–multimethod matrices in consumer research. *Journal of Consumer Research, 17*, 426–439.

Bandura, A. (1984). *Social foundations of thought and action: A social cognitive theory*. Englewood cliffs, NJ: Prentice-Hall

Bandura, A. (1990). Mechanisms of moral disengagement. In W. Reich (Ed.), *Origins of terrorism: Psychologies, ideologies, theologies, states of mind,* 161-191. Cambridge: Cambridge University Press.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology, 3*(3)*,* 193-209.

Bandura, A., Barbaranelli, C., Caprara, G., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology. 71*, 364-374.

Barriga, A. Q., & Gibbs, J. C. (1996). Measuring cognitive distortion in antisocial youth: Development and preliminary validation of the “How I Think” questionnaire. *Aggressive Behavior, 22*, 333-343.

Barriga, A. Q., Landau, J. R., Stinson, B. L., Liau, A. K., & Gibbs, J. C. (2000). Cognitive distortions and problem behaviors in adolescents. *Criminal Justice and Behavior, 27*(1), 36-56

*Barsky, A. (2011). Investigating the effects of moral disengagement and participation on unethical work behavior. *Journal of Business Ethics, 104*(1), 59-75.

Bennet, R. J., & Robinson, S. L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology, 85*(3), 349-360.

Benson, M. L. (1985). Denying the guilty mind: Accounting for involvement in white-collar crime. *Criminology 23,* 583-608.

Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology, 92*(2), 410-424.

Berry, C. M., Sackett, P. R., & Wiemann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology, 60*, 271-301.

*Betts, T. K., Setterstrom, A. J., Pearson, J. M., & Totty, S. Explaining cyberloafing through a theoretical integration of theory of interpersonal behavior and theory of organizational justice. *Journal of Organizational and End User Computing, 26*(4), 23-42.

Bibring, G.L., Dwyer, T.F., Huntington, D.S., & Valenstein, A.F. (1961). A study of the psychological processes in pregnancy and of the earliest mothe-rchild relationship. II. Methodological considerations. *Psychoanalytic Study of the Child, 16*, 25-72.

Blau, P. M. (1964). *Exchange and power in social life*. New York: Wiley.

Bone, H. (1975). The inter-personal and the intra-personal. *Contemporary Psychoanalysis, 11*, 46-65.

*Bonner, J. M., Greenbaum, R. L., & Mayer, D. M. (2016). My boss is morally disengaged: The role of ethical leadership in explaining the interactive effect of supervisor and employee moral disengagement on employee behaviors. *Journal of Business Ethics, 137*(4), 731–742.

Bowling, N. A., Gibson, A. N., & DeSimone, J. A. (2022). Stop with the questions already! Does data quality suffer for scales positioned near the end of a lengthy questionnaire? *Journal of Business and Psychology*. https://doi.org/10.1007/s10869-021-09787-8

Campbell, D. T., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Bryant, E., Schmike, E. B., Brehm, H. N., & Uggen, C. (2017). Techniques of Neutralization and identify work among accused genocide perpetrators. *Social Problems, 1*, 1-19.

*Chen, M., Chen, C. C., & Sheldon, O. J. (2016). Relaxing moral reasoning to win: How organizational identification relates to unethical pro-organizational behavior. *Journal of Applied Psychology, 101*(8), 1082-1096.

*Chen, H., Chau, P. Y. K., & Li, W. (2018). The effects of moral disengagement and organizational ethical climate on insiders' information security policy violation behavior. *Inf. Technol. People, 32*, 973-992.

*Chen, M., & Liang, J. (2017). High performance expectation and unethical pro-organizational behavior: Social cognitive perspective. *Acta Psychologica Sinica, 49*(1), 94-105.

*Cheng, L., Li, W., Zhai, Q., & Smyth, R. (2014). Understanding personal use of the internet at work: An integrated model of neutralization techniques and general deterrence theory. *Computers in Human Behavior, 38*, 220-228

*Christian, J. S., & Ellis, A. P. J. (2014). The crucial role of turnover intentions in transforming moral disengagement into deviant behavior at work. *Journal of Business Ethics, 119*, 193-208.

*Claybourn, M. (2011). Relationships between moral disengagement, work characteristics, and workplace harassment. *Journal of Business Ethics*, *100*, 283-301.

Coleman. J. (1994). *The Criminal Elite: The Sociology of White Collar Crime*. New York: St. Martins.

Comrey, A. L., & Lee, H. B. (1992). A first course in factor analysis. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cromwell, P. & Thurman, Q. (2003). The devil made me do it: Use of neutralizations by shoplifters. *Deviant Behavior 24*(6), 535-550.

Cuadrado, D., Salgado, J.F., Moscoso, S. (2019). Prevalence and Correlates of Academic Dishonesty: Towards a Sustainable University. *Sustainability, 11*, 6062.

Cuadrado, D., Salgado, J.F., Moscoso, S. (2020) Individual differences and counterproductive academic behaviors in high school. *PLoS ONE, 15*.

Cuadrado, D., Salgado, J. F., & Moscoso, S. (2021). Personality, intelligence, and counterproductive academic behaviors: A meta-analysis. *Journal of Personality and Social Psychology, 120*(2), 504-537.

Dahlke, J. A., & Wiernik, B. M. (2018). psychmeta: An R package for psychometric meta-analysis. *Applied Psychological Measurement.*

Dalal, R. S. (2005) A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *Journal of applied Psychology, 90*(6). 1241-1255.

*Dang, C. T., Umphress, E. E., & Mitchell, M. S. (2017). Leader social accounts of subordinates' unethical behavior: Examining observer reactions to leader social accounts with moral disengagement language. *Journal of Applied Psychology, 102*(10), 1448-1461.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*(5), 880-896.

Dilchert, S., Ones, D. S., Davis, R. D., & Rostow, C. D. (2007). Cognitive ability predicts objectively measured counterproductive work behaviors. *Journal of Applied Psychology, 92,* 616-627.

*Duffy, M. K., Scott, K. L., Shaw, J. T., Tepper, B. J., & Aquino, K. (2012). A social context model of envy and social undermining. *The Academic of Management Journal, 55*(3), 643-666.

Dunford, F. W. & Kunz, P. R. 1973. The neutralization of religious dissonance. *Review of Religious Research 15*, 2-9.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56* ,455–463.

Ellen, B. P. III, Alexander, K. C., Mackey, J. D., McAllister, C. P., & Carson, J. E. (2021). Portrait of a workplace deviant: A clearer picture of the Big Five and Dark Triad as predictors of workplace deviance. *Journal of Applied Psychology, 106*(12), 1950–1961.

*Erkutlu, H., & Chafra, J. B. (2019). Leader psychopathy and organizational deviance: The mediating role of psychological safety and the moderating role of moral disengagement. *International Journal of Workplace Health Management*.

Everitt, B. S. (1975). Multivariate analysis: The need for data, and other problems. British Journal of Psychiatry, 126, 237–240.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 56-83

*Fehr, R., Fulmer, A., & Keng-Highberger, F. T. (2019). How do employees react to leaders' unethical behavior? The role of moral disengagement. *Personnel Psychology, 73*(1), 73-93.

*Fehr, R., Welsh, D., Yam, K. C., Baer, M., Wei, W., & Vaulont, M. (2019). The role of moral decoupling in the causes and consequences of unethical pro-organizational behavior. *Organizational Behavior and Human Decision Processes, 153*, 27-40.

*Fida, R, Paciello M., Tramontano, C., Fontaine, RG Barbaranelli, C., & Farnese, ML. (2015). An integrative approach to understanding counterproductive work behavior: The roles of stressors, negative emotion and moral disengagement. *Journal of business ethics, 130*, 131-144.

*Fida, R., Tramontano, C., Paciello, M., Guglielmetti, C., Gilardi, S., Probst, T. M., & Barbaranelli, C. (2018). ‘First, do no harm’: The role of negative emotions and moral disengagement in understanding the relationship between workplace aggression and misbehavior. *Frontiers in Psychology, 9,* 1-17.

Fishbein, M., & Ajzen, I. (1974). Attitudes towards objects as predictors of single and multiple behavioral criteria. *Psychological Review* , *81* , 59–74. http://dx.doi.org/10.1037/h0035872

Fraley, R. C., & Vazire, S. (2014). The N-pact factor: evaluating the quality of empirical journals with respect to sample size and statistical power. *PloS One, 9*(10), e109019. doi:10.1371/journal.pone.0109019

Freud, A. (1981). Insight - its presence and absence as a factor in normal development. *Psychoanalytic Study of the Child, 36*, 241-9.

Freud, S (1901a). *On dreams*. SE 5: 63186.

Freud, S. (1914). *On the history of the psycho-analytic movement*. SE 14: 766.

Frobinger, M. Q. (2011). Using one's eyeballs to evaluate cumulative meta-analyses: A test for publication bias analyses. *Journal of Applied Sightology, 100*, 1-365.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*, 156–168.

*Gan C. (2018). Ethical leadership and unethical employee behavior: A moderated mediation model. *Social Behavioral and Personality, 46*(8), 1271-1284.

Garrett, D., Bardford, J., Meyers. R., & Becker, J. (1989). Issues management and organizational accounts: An analysis of corporate responses to accusations of unethical business practices. *Journal of Business Ethics, 8*(7), 507-520.

Gignac, E. G., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74-78.

Giluk, T. L., & Postlethwaite, B. E. (2014). Big Five personality and academic dishonesty: A meta-analytic review. *Personality and Individual Differences, 72*, 59–67.

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Guttman, L. (1955). The determinacy of factor score matrices with applications for five other problems of common factor theory. *British Journal of Statistical Psychology, 8*, 65–82.

Haines, V. J., Diekhoff, G. M., LaBeff, E. E., & Clark, R. E. (1986). College cheating: Immaturity, lack of commitment, and the neutralizing attitude. *Research in Higher Education, 25, 342–354.*

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). Multivariate data analysis (4th ed.). Saddle River, NJ: Prentice Hall.

Harding, T. S., Carpenter, D. D., Finelli, C. J., & Passow, H. J. (2004). *The influence of academic dishonesty on ethical decision-making in the workplace: A study of engineering students*. Proceedings of ASEE 111th annual conference and exposition, Salt Lake City, UT. https://doi.org/10.18260/1-2–13993

Harding, T. S., Finelli, C. J., & Carpenter, D. D. (2006). *Cheating in college and its influence on ethical behavior in professional engineering practice*. Proceedings of ASEE 113th annual conference & exposition, Chicago, IL. https://doi.org/10.18260/1-2–357

*Harris, L. C., & He, H. (2019). Retail employee pilferage: A study of moral disengagement. *Journal of Business Research, 99*, 57-68.

*He, P., Peng, Z., Zhao, H., & Estay, C. (2017). How and when compulsory citizenship behavior leads to employee silence: A moderated mediation model based on moral disengagement and supervisor-subordinate guanxi views. *Journal of Business Ethics, 155*, 259-274.

Henle, C. A. (2005). Predicting workplace deviance from the interaction between organizational justice and personality. *Journal of Managerial Issues, 17*, 247−263.

Henry, S., & Eaton, R. (1994). *Degrees of Deviance: Student Accounts of Their Deviant Behavior*. Salem: Sheffield Publishing Company.

Hessen, P., Lee, S. D., & Kuncel, N. R. (2022). *Assessment of social and emotional learning for admissions and selection*. In Assessing Competencies for Social and Emotional Learning (pp. 169-188). Routledge.

Higgins J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane Handbook for Systematic Reviews of Interventions* Version 6.0. The Cochrane Collaboration. Available from www.handbook.cochrane.org.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ : British Medical Journal*, *327*(7414), 557–560.

Hooper, D., Coughlan, J., & Mullen, M. R. (2007). Structural equation modeling: Guidelines for determining model fit. *Electronic Journal on Business Research Methods, 6*(1), 53-60.

Howard, J. A., & Levinson, R. (1985). The overdue courtship of attribution and labeling. *Social Psychology Quarterly,* 191-202.

Hu, L & Benter, P. M. (1999). Cutoff criterio for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

*Huang, Y., Greenbaum, R. L., Bonner, J. M., & Wang, C. S. (2018). Why sabotage customers who mistreat you? Activated hostility and subsequent devaluation of targets as a moral disengagement mechanism. *Journal of Applied Psychology, 104*(4), 495-510.

*Huang, G., Wellman, N., Ashford, S. J., Lee, C., & Wang, L. (2017). Deviance and exit: The organizational costs of job insecurity and moral disengagement. *Journal of Applied Psychology, 102*(1), 26–42.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: correcting error and bias in research findings (2nd ed.).* Thousand Oaks, CA: Sage Publications, Inc.

Hurtz G. M. & Alliger G. M. (2002). Influence of coaching on integrity test performance and unlikely virtue scale scores. *Human Performance, 15*, 255–273.

*Hystad, S. W., Mearns, K. J., & Eid, J. (2014). Moral disengagement as a mechanism between perceptions of organizational injustice and deviant work behaviours. *Safety Science, 68*, 138-145.

International Cognitive Ability Resource Team (2014). International cognitive ability resource. Retrieved from https://icar-project.com/

Inzlicht, M., Gervias, W., & Berkman, E. (2015). *Bias-correction tech-niques alone cannot determine whether ego depletion is different from zero: Commentary on Carter, Kofler, Forster, & McCullough, 2015*. Unpublished manuscript. http://dx.doi.org/10.2139/ssrn.2659409

Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials, 4*, 245–253.

Jacobson, J. (2020). *The Effect of Stress on the Neutralization of Misconduct.* [Unpublished bachelor thesis]. University of Minnesota, Minneapolis, Minnesota.

Jones, D. N. (2009). *The personality of a cheater*. Unpublished Manuscript.

Jones, D. N., Olderbak, S. G., & Figueredo, A. J. (2011). The intentions towards infidelity scale. *Handbook of sexuality-related measures*, 251-253.

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment, 21*(1), 28-41. https://doi.org/10.1177/1073191113514105

*Kacmar, K. M., Carlson, D. S., Thompson, M. & Zivnuska, S. (2019). But I still feel guilt: A test of a moral disengagement propensity model. *Human Performance, 32*, 165-180.

Kaptein, M. & van Helvoort, M. (2019). A model of neutralization techniques. *Deviant Behavior, 40*(10), 1260-1285. https://doi.org/10.1080/01639625.2018.1491696

*Keem, S., Shalley, C. E., Kim, E., & Jeong, I. (2018). Are creative individuals bad apples? A dual pathway model of unethical behavior. *Journal of Applied Psychology, 103*(4). 416-431.

*Khansa, L., Kuem, J., Siponen, M., & Kim, S. S. (2017). To cyberloaf or not to cyberloaf: The impact of the announcement of formal organizational controls. *Journal of Management Information Systems, 34*(1), 141-176.

Klenowski, P. M., Copes, H., Mullins, C. W. (2011). Gender, identity and accounts: How white-collar offenders do gender when making sense of their crimes. *Justice Quarterly, 28*, 46–69.

Kline, P. (1994). An easy guide to factor analysis. New York: Routledge.

Klockars, C. (1974). *The Professional Fence.* New York: Free Press.

Knoll, M., Lord, R. G., Petersen, L. E., & Weigelt, O. (2016). Examining the moral grey zone: The role of moral disengagement, authenticity, and situational strength in predicting unethical managerial behavior. *Journal of Applied Social Psychology, 46*, 65–78.

*Kong, M., & Yuan, Y. (2018). Perceived deviance tolerance: Make employees' moral constructs accessible from a dual-system. *Management Decision, 56,* 1936-1955.

Kuncel, N. R. & Tellegen, A. (2009). A conceptual and empirical reexamination of the measurement of social desirability items: Implications for detecting desirable response style and scale development. *Personnel Psychology, 62*(2), 201-228.

Kvalnes Ø. (2019) Moral Neutralization. In: *Moral Reasoning at Work*. Palgrave Pivot, Cham. https://doi.org/10.1007/978-3-030-15191-1_13

*Lee, A., Schwarz, G., Newman, A., & Legood, A. (2017). Investigating when and why psychological entitlement predicts unethical pro-organizational behavior. *Journal of Business Ethics, 154*, 109-126.

*Lee, K., Kim, E., Bhave, D. P., & Duffy, M. K. (2016). Why victims of undermining at work become perpetrators of undermining: An integrative model. *Journal of Applied Psychology, 101*(6), 915-924.

Lee, S.D. & Kuncel, N. R. (2020). *Neutralization, the big five, and workplace outcomes.* Poster presented at the American Psychological Association Annual Conference.

Lee, S. D., Kuncel, N. R., & Gau, J. (2020). Personality, attitude, and demographic correlates of academic dishonesty: A meta-analysis. *Psychological Bulletin, 146*(11), 1042-1058.

Levi, K. (1981). Becoming a hit man: Neutralization in a very deviant career. *Urban Life, 10,* 47-63.

Levin, S. (1978). Internal mechanisms that support the acting out of destructive impulses. *International Review of Psychoanalysis, 5*, 55-60.

*Lim, V. K. G. (2002). The IT way of loafing on the job: Cyberloafing, neutralizing, and organizational justice. *Journal of Organizational Behavior, 23(5),* 675-694.

Lișman, C.G., & Holman, A.C. (in press). Innocent cheaters: a new scale measuring the moral disengagement of marital infidelity. *Studia Psychologica, 64*(2).

Lowry, P. B., Zhang, J., Wang, C., & Siponen, M. (2016). Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning model. *Information Systems Research, 27*(4), 962-986.

Marcus, B., Ashton, M. C., & Lee, K. (2013). A note on the incremental validity of integrity tests beyond standard personality inventories for the criterion of counterproductive behavior. *Canadian Journal of Administrative Sciences, 30*, 18-25

Maruana, S., & Copes, H. 2005. What have we learned from five decades of neutralization research? in *Crime and Justice: A Review of Research (Volume 32)*. Chicago: The University of Chicago Press, 221-320.

Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2018). Assessing Fit in Structural Equation Models: A Monte-Carlo Evaluation of RMSEA Versus SRMR Confidence Intervals and Tests of Close Fit. Structural Equation Modeling: A Multidisciplinary Journal, 25(3), 389–402. http://doi.org/10.1080/10705511.2017.1389611

McCabe, D. L., & Trevino, L. K. (1993). Academic dishonesty: Honor codes and other contextual influences. *The Journal of Higher Education, 64*, 522–538.

McCarthy, J. G., & Stewart, A. L. (1998). Neutralisation as a process of graduated desensitisation: Moral values of offenders. *International Journal of Offender Therapy and Comparative Criminology, 42*(4), 278-290.

McDaniel, M. A. (2009) Cumulative meta-analyses as a publication bias method. *Paper presented at the 24th Annual Meeting of the Society for Industrial and Organizational Psychology*. New Orleans

McDonald, R. P. (1999). Test Theory: A Unified Treatment. Mahwah, NJ: Erlbaum.

*McFerran, B., Aquino, K., & Duffy, M. (2010). How personality and moral identity relate to individuals' ethical ideology. *Business Ethics Quarterly, 20*(1), 35-56.

Meng, C. L., Othman, J., D'Silva, J. L., Omar, Z. (2014). Influence of neutralization attitude in academic dishonesty among undergraduates. *International Education Studies*, 7(6), 66-73.

Minor, W. 1981. Techniques of neutralization: A reconceptualization and empirical examination. *Journal of Research in Crime and Delinquency, 18*(2), 295-318.

*Moore, C., Detert, J. R., Trevino, L. K., Baker, V. L., & Mayer, D. M. (2012). Why employees do bad things: Moral disengagement and unethical organizational behavior. *Personnel Psychology, 65,* 1-48.

*Moore, C., Mayer, D. M., Chiang, F. F. T., Crossley, C., Karlesky, M. J., & Birtch, T. A. (2019). Leaders matter morally: The role of ethical leadership in shaping employee moral cognition and misconduct. *Journal of Applied Psychology, 104*(1), 123–145. https://doi.org/10.1037/apl0000341

Moreno, S. G., Sutton, A. J., Ades, A. E., Stanley, T. D., Abrams, K. R., Peters, J. L., et al.

(2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *Biomedical Medical Research Methodology*. 9, 1–17. https://doi.org/10.1186/1471-2288-9-2

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing, 5*(2), 159-168.

Newman, A., Le, H., North-Samardzic, A., & Cohen, M. (2019). Moral disengagement at work: A review and research agenda. *Journal of Business Ethics*. https://10.1007/s1055101904173-0

*Nguyen, B. (2015). *The effect of ethical leadership, behavioural integrity, and moral disengagement in predicting turnover intentions during newcomer socialization* [Unpublished doctoral dissertation]. University of Calgary.

*Niederhauser, C. (2019). *Unethical pro-organizational behavior and organizational identification: Exploring the moderating effects of conscientiousness and neuroticism* [Unpublished master's thesis]. Illinois State University.

*Ntayi, J., Eyaa, S., & Ngoma, M. (2010). Moral disengagement and the social construction of procurement officers' deviant behaviours.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill.

O'Boyle, E. H., Jr., Forsyth, D. R., Banks, G. C., & McDaniel, M. A. (2012). A meta-analysis of the Dark Triad and work behavior: A social exchange perspective. *Journal of Applied*

Psychology, 97(3), 557-579. http://dx.doi.org.ezp1.lib.umn.edu/10.1037/a0025679

Ogunfowora, B. T., Nguyen, V. Q., Steel, P., & Hwang, C. C. (2022). A meta-analytic investigation of the antecedents, theoretical correlates, and consequences of moral disengagement at work. *Journal of Applied Psychology, 107*(5), 746.

Ones DS. (1993). *The construct validity of integrity tests*. Unpublished doctoral dissertation, University of Iowa, Iowa city, Iowa.

Ones, D. S., & Dilchert, S. (2013). Counterproductive work behaviors: Concepts, measurement, and nomological network. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology* (Vol. 1: Test theory and testing and assessment in industrial and organizational psychology, pp. 643–659). Washington, DC: American Psychological Association.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of applied psychology, 81*(6), 660.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*(4), 679–703.

Organ, D. W. (2018). *The roots of organizational citizenship behavior.* In Podsakoff, P. M., Mackenzie, S. B., & Podsakoff, N. P. The Oxford Handbook of Organizational Citizenship Behavior

*Paciello, M., Fida, R., Tramontano, C., Ghezzi, V., & Barbaranelli, C. (2019). Phenomenological configurations of workplace bullying: A cluster approach. *Personality and Individual Differences, 151*, 1-6.

*Pagliaro, S., Presto, A. L., Barattucci, M., Giannella, V. A., & Barreto, M. (2018). On the effects of ethical climate(s) on employees' behavior: A social identity approach. *Frontiers in Psychology, 9,* 1-10.

Paula, J. J., de Souza Costa, D., de Miranda, D. M., & Romano-Silva, M. A. (2020). The abbreviated version of the Barratt Impulsiveness Scale (ABIS): Psychometric analysis, reliable change indexes in clinical practice and normative data. *Psychiatry Research, 291,* 113120.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Eds.), Measures of personality and social psychological attitudes, 17-59. San Diego, CA: Academic Press.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., and Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine. 26*, 4544–4562.

*Pettita, L., Probst, T. M., & Barbaranelli, C. (2017). Safety culture, moral disengagement, and accident underreporting. *Journal of Business Ethics, 141,* 489-504.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology, 88*(5), 879.

*Probst, T. M., Petitta, L., Barbaranelli, C., & Austin, C. (2020). Safety-related moral disengagement in response to job insecurity: Counterintuitive effects of perceived organizational and supervisor support. *Journal of Business Ethics, 162*, 343-358.

*Qin, X., Dust, S. B., DiRenzo, M. S., & Wang, S. (2019). Negative creativity in leader-follower relations: A daily investigation of leaders' creative mindset, moral disengagement, and abusive supervision. *Journal of Business Psychology, 35,* 665-682.

Ree, M. J., & Earles, J. A. (1993). g is to psychology what carbon is to chemistry: A reply to Sternberg and Wagner, McClelland, and Calfee. *Current Directions in Psychological Science*, *2*(1), 11-12. https://dx.doi.org/10.1111/1467-8721.ep10770509

*Reynolds, S. J., Dang, C. T., Yam, K. C., & Leavitt, K. (2014). The role of moral knowledge in everyday immorality: What does it matter if I know what is right? *Organizational Behavior and Human Decision Processes, 123*, 124-137.

Ribeaud, D. & Eisner, M. (2010). A moral disengagement, neutralization techniques, and self-serving cognitive distortions the same? Developing a unified scale of moral neutralization of aggression. *International Journal of Conflict and Violence, 4(2),* 298 -315.

Rodriguez, A., Reise, S. P., and Haviland, M. G. (2016). Evaluating bifactor models: calculating and interpreting statistical indices. Psychol. Methods 21, 137–150.

Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of applied psychology, 87*(1), 66.

Rücker, G., Carpenter, J.R. and Schwarzer, G. (2011), Detecting and adjusting for small-study effects in meta-analysis. *Biometrical Journal, 53.* 351-368.

Ryan, A. M. & Sackett, P. R. (1987). Pre-employment honesty testing: Fakability, reactions of test takers, and company image. *Journal of Business and Psychology, 1*(3), 248-256.

Sackett, P. R., Burris, L. R., & Callahan, C. (1989). Integrity testing for personnel selection: An update. *Personnel Psychology, 42*, 491-529.

Sackett, P. R., & DeVore, C. J. (2001). Counterproductive behaviors at work. In. N. Anderson, DS Ones, HK Sinangil, & C. Viswesvaran (Eds.). *Handbook of industrial, work and organizational psychology,* 145-164.

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2021). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*.

*Sahi, Q. B., & Ahmad, M. (2019). Impact of job insecurity and moral disengagement on counterproductive work behavior. *City University Research Journal, 9*(2), 279-294.

*Saidon, I. M. (2012). *Moral disengagement in manufacturing: A Malaysian study of* *antecedents and outcomes* [Unpublished doctoral dissertation]. Curtin University.

*Samnami, A., Salamon, S. D., & Singh, P. (2014). Negative affect and counterproductive workplace behavior: The moderating role of moral disengagement and gender. *Journal of Business Ethics, 119*, 235-244.

Sandler, J. (1976). Dreams, unconscious fantasies and identity of perception. International *Review of Psychoanalysis, 3*, 33-42.

Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta‐analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology, 65*(3), 445-494.

Shields, I. W., & Whitehall, G. C. (1994). Neutralization and delinquency among teenagers. *Criminal Justice and Behavior, 21*(2), 223-235.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2). 262-274.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62(5), 529-540.

*Shaw, K., & Liao, H. (2018). How does authoritarian leadership lead to employee unethical pro-organizational behavior? The mediating effect of work stressor and moral disengagement. *Advances in Economics, Business and Management Research, 51*, 86-94.

Shields, I. W. & Whitehall, G., C. (1994). Neutralization and delinquency among teenagers. *Criminal Justice and Behavior 21*(2), 223-35.

Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin, 37*(3), 330-349

Schulze, R. (2004). *Meta-analysis*: A comparison of approaches. Hogrefe & Huber Publishers.

*Seriki, O.K., Nath, P., Ingene, C., & Evans, K. (2020). How complexity impacts salesperson counterproductive behavior: The mediating role of moral disengagement. *Journal of Business Research, 107*, 324-335.

Stanley, T.D. (2005), Beyond Publication Bias. *Journal of Economic Surveys, 19*. 309-345.

Sykes, G. M. & Matza, D. (1957). Techniques of neutralization: A theory of delinquency. *American Sociological Review, 22*(6). 664-670.

Swann, A., Lennon, A., & Cleary, J. (2017). *Driving Moral Disengagement Scale* [Database record]. Retrieved from PsycTESTS.

*Teh, P., Ahmed, P. K., & D'Arcy, J. (2015). What drives information security policy violations among banking employees? Insights from neutralization and social exchange theory. *Journal of Global Information Management, 23*(1), 44-64.

Terrin, N., Schmid, C. H., Lau, J., and Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine. 22*, 2113–2126.

Thurman, Q. C. (1984). Deviance and the neutralization of moral commitment: A empirical analysis. *Deviant Behavior, 5*, 291-304.

Umphress, E. E. & Bingham, J. B. (2011). When employees do bad things for good reasons: Examining unethical pro-organizational behaviors. *Organizational Science, 23*(3). 621-640.

*Umphress, E. E., Gardner, R. G., Stoverink, A. C., & Leavitt, K. (2019). Feeling activated and acting unethically: The influence of activated mood on unethical behavior to benefit a teammate. *Personnel Psychology, 73*(1), 95-123.

*Valle, M., Kacmar, K. M., & Zivnuska, S. (2019). Understanding the effects of political environments on unethical behavior in organizations. *Journal of Business Ethics, 156*, 173-188.

*Valle, M, Kacmar, K. M., Zivnuska, S., & Harting, T. (2019). Abusive supervision, leader-member exchange, and moral disengagement: A moderated-mediation model of organizational deviance, *The Journal of Social Psychology, 159*(3), 299-312.

Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology, 97*(3), 499-530. doi:10.1037/a0021196

Veris Benchmarks. (2023). *Efficacy of Testing*. Verishbenchmarks.com/efficacy-of-testing/

Viano, E. C., Jain, G., & Sahni, S. P. (2016). Understanding attribution bias and reasons behind internet infidelity in India. *Cybercrime, Organized Crime, & Societal Responses*, 47-65.

Viechtbauer, W. (2005). Bias and Efficiency of Meta-Analytic Variance Estimators in the

Random-Effects Model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261–293.

Viechtbauer W (2010). "Conducting meta-analyses in R with the metafor package." Journal of *Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Viswesvaran, C. & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement 59*(2), 197–210.

Wanek, J. E., Sackett, P. R., & Ones, D. S. (2003). Towards an understanding of integrity test similarities and differences: An item‐level analysis of seven tests. *Personnel Psychology, 56*(4), 873-894.

Whitener, E. M (1990) Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology*, 75(3), 315-321.

Whitley, B. E., Jr. (1998). Factors associated with cheating among college students: A review. *Research in Higher Education, 39,* 235–274. http://dx.doi.org/10.1023/A:1018724900565

Willott, S., Griffin, C., Torrance, M. (2001). Snakes and ladders: Upper-middle class male offenders talk about economic crime. *Criminology, 39*, 441–466.

Zepf, S. (2011). About rationalization and intellectualization. *International Forum of Psychoanalysis, 20*(3), 148-158. http://dx.doi.org/10.1080/0803706X.2010.550316

*Zhang, J., Akhtar, M. N., Zhang, Y., & Sun, S. (2019). Are overqualified employees bad apples? A dual-pathway model of cyberloafing. *Internet Research.*

*Zhang, P., Li, S., Liu, W., Han, Y., Muhammad, N. A. (2018). Exploring the role of moral disengagement in the link between perceived narcissistic supervision and employees' organizational deviance: A moderated mediation model. *Asian Journal of Social Psychology, 21*, 223-236.

*Zheng, X., Qin, X., Liu, X., & Liao, H. (2019). Will creative employees always make trouble? Investigating the roles of moral identity and moral disengagement. *Journal of Business Ethics*, *157*(3), 653-672.

References marked with * were included in the Study 2 meta-analysis

Appendices

Appendix 1

Haines et al. (1986) measure of neutralization

"Jack should not be blamed for cheating if..."

Q128 the course material is too hard

- Strongly agree (1)
- Agree (2)
- Neutral (3)
- Disagree (4)
- Strongly disagree (5)

Q129 he is in danger of losing his scholarship

- Strongly agree (1)
- Agree (2)
- Neutral (3)
- Disagree (4)
- Strongly disagree (5)

Q130 he doesn't have time to study

- Strongly agree (1)
- Agree (2)
- Neutral (3)
- Disagree (4)
- Strongly disagree (5)

Q131 the instructor doesn't seem to care

- Strongly agree (1)
- Agree (2)
- Neutral (3)
- Disagree (4)
- Strongly disagree (5)

Q132 the instructor acts like his/her course is the only one

- Strongly agree (1)
- Agree (2)
- Neutral (3)
- Disagree (4)
- Strongly disagree (5)

Q133 his cheating isn't hurting anyone

- Strongly agree (1)
- Agree (2)
- Neutral (3)
- Disagree (4)
- Strongly disagree (5)

Q134 everyone else in the room seems to be cheating

- ◯ Strongly agree (1)
- ◯ Agree (2)
- ◯ Neutral (3)
- ◯ Disagree (4)
- ◯ Strongly disagree (5)

Q135 the people sitting around him made no attempt to cover their papers

- ◯ Strongly agree (1)
- ◯ Agree (2)
- ◯ Neutral (3)
- ◯ Disagree (4)
- ◯ Strongly disagree (5)

Q136 his friend asked him to help him/her cheat

- Strongly agree (1)
- Agree (2)
- Neutral (3)
- Disagree (4)
- Strongly disagree (5)

Q137 the instructor left the room

- Strongly agree (1)
- Agree (2)
- Neutral (3)
- Disagree (4)
- Strongly disagree (5)

Q138 the course is required

- ○ Strongly agree (1)
- ○ Agree (2)
- ○ Neutral (3)
- ○ Disagree (4)
- ○ Strongly disagree (5)

Appendix 2

Thurman (1984) measure of neutralization

Q120 It is not wrong to violate the law when the victim involved is a dishonest person.

◯ Strongly disagree (1)

◯ Disagree (2)

◯ Agree (3)

◯ Strongly agree (4)

Q121 It is all right to break the law if it is done to aid a friend in need.

◯ Strongly disagree (1)

◯ Disagree (2)

◯ Agree (3)

◯ Strongly agree (4)

Q122 It is okay to break the law if you aren't sure what the law is.

◯ Strongly disagree (1)

◯ Disagree (2)

◯ Agree (3)

◯ Strongly agree (4)

Q123 It is okay to break the law if no one gets hurt.

- Strongly disagree (1)
- Disagree (2)
- Agree (3)
- Strongly agree (4)

Q124 It is all right to break the law under circumstances where it seems like you have little other choice.

- Strongly disagree (1)
- Disagree (2)
- Agree (3)
- Strongly agree (4)

Q125 It is more acceptable for an honest and law-abiding citizen to break the law than it would be for a frequently dishonest person to do so.

- Strongly disagree (1)
- Disagree (2)
- Agree (3)
- Strongly agree (4)

Q126 It is not as wrong to break laws which seem unfair and unjust to you.

○ Strongly disagree (1)

○ Disagree (2)

○ Agree (3)

○ Strongly disagree (4)

Appendix 3

Agnew and Peters (1986) measure of neutralization

Would you feel guilty for shoplifting if:

Q140
The store owner's insurance covered the loss?

- Very guilty (1)
- Somewhat guilty (2)
- Not guilty (3)

Q141 You needed the item very badly but did not have the money to pay for it?

- Very guilty (1)
- Somewhat guilty (2)
- Not guilty (3)

Q142 The store owner was dishonest and often cheated his customers?

- Very guilty (1)
- Somewhat guilty (2)
- Not guilty (3)

Q143 The store owner was wealthy and wouldn't miss the item?

- ◯ Very guilty (1)
- ◯ Somewhat guilty (2)
- ◯ Not guilty (3)

Q144 Many of your friends shoplifted?

- ◯ Very guilty (1)
- ◯ Somewhat guilty (2)
- ◯ Not guilty (3)

Appendix 4

Swan et al. (2017) Measure of Moral Disengagement

Q146
It's okay to yell at other drivers who put the lives of your passengers at risk.

- ○ Strongly disagree (1)
- ○ Disagree (2)
- ○ Neutral (3)
- ○ Agree (4)
- ○ Strongly Agree (5)

Q147 Honking the horn loudly is just a way of letting off frustration.

- ○ Strongly disagree (1)
- ○ Disagree (2)
- ○ Neutral (3)
- ○ Agree (4)
- ○ Strongly agree (5)

Q148 Yelling at other drivers is pretty tame when compared to people that attack other drivers.

- Strongly disagree (1)
- Disagree (2)
- Neutral (3)
- Agree (4)
- Strongly agree (5)

Q149 Speeding a little over the limit is not too serious compared to those that speed a lot over the limit.

- Strongly disagree (1)
- Disagree (2)
- Neutral (3)
- Agree (4)
- Strongly agree (5)

Q150 If a driver is pushed into being rude to other drivers they shouldn't be blamed for it.

- Strongly disagree (1)
- Disagree (2)
- Neutral (3)
- Agree (4)
- Strongly agree (5)

Q151 It's ok to go over the speed limit if it means you are keeping up with the rest of traffic.

- Strongly disagree (1)
- Disagree (2)
- Neutral (3)
- Agree (4)
- Strongly agree (5)

Q152 Drivers don't mind being honked at because they know it just means 'hurry up'.

- ◯ Strongly disagree (1)
- ◯ Disagree (2)
- ◯ Neutral (3)
- ◯ Agree (4)
- ◯ Strongly agree (5)

Q153 Flashing headlights to get someone to move over, doesn't really hurt anyone.

- ◯ Strongly disagree (1)
- ◯ Disagree (2)
- ◯ Neutral (3)
- ◯ Agree (4)
- ◯ Strongly agree (5)

Q154 If you are getting honked at while driving you probably deserve it.

- Strongly disagree (1)
- Disagree (2)
- Neutral (3)
- Agree (4)
- Strongly agree (5)

Q155 People who don't know how to drive, provoke bad driving in others.

- Strongly disagree (1)
- Disagree (2)
- Neutral (3)
- Agree (4)
- Strongly agree (5)

Q156 It's alright to abuse drivers who are behaving like "knobs".

- Strongly disagree (1)
- Disagree (2)
- Neutral (3)
- Agree (4)
- Strongly agree (5)

Q157 A driver who is inconsiderate doesn't deserve to be treated like a normal person.

- Strongly disagree (1)
- Disagree (2)
- Neutral (3)
- Agree (4)
- Strongly agree (5)

Q158 Some drivers deserve to be treated like the idiots they are.

- ◯ Strongly disagree (1)
- ◯ Disagree (2)
- ◯ Neutral (3)
- ◯ Agree (4)
- ◯ Strongly agree (5)

Appendix 5

Moore et al. (2012) Measure of Moral Disengagement

Q97
It is okay to spread rumors to defend those you care about.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

Q98 Taking something without the owner’s permission is okay as long as you’re just borrowing it.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

Q99 Considering the ways people grossly misrepresent themselves, it’s hardly a sin to inflate your own credentials a bit.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

Q100 People shouldn't be held accountable for doing questionable things when they were just doing what an authority figure told them to do.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

Q101 People can't be blamed for doing things that are technically wrong when all their friends are doing it too.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

Q102 Taking personal credit for ideas that were not your own is no big deal.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

Q103 Some people have to be treated roughly because they lack feelings that can be hurt.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

Q104 People who get mistreated have usually done something to bring it on themselves.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

Appendix 6

Descriptive Statistics of the Total Sample

Source	N
MTurk	375
REP	328
Prolific	738

Total Sample

Variable	N	Mean	Std. Dev.	Min	Max	Kurtosis	Skew
g	1404	-0.006	0.925	-1.936	2.909	0.06	0.53
S_1	1404	-0.003	0.821	-2.427	3.46	0.64	0.59
S_2	1404	-0.002	0.869	-2.627	3.187	0.28	0.1
S_3	1404	-0.007	0.95	-1.985	4.545	5.86	1.86
S_4	1404	-0.01	0.729	-2.679	3.071	0.27	0.14
S_5	1404	0.001	0.853	-2.361	2.975	-0.03	0.49
MD_Comp	1441	0	0.888	-1.871	3.044	-0.11	0.15

Neut_Comp	**1437**	**0.001**	**0.801**	**-1.274**	**2.566**	**-0.42**	**0.36**
MNeut_Comp	**1437**	**0.001**	**0.763**	**-1.513**	**2.485**	**-0.24**	**0.22**
T.Moore	**1441**	**2.5**	**1**	**1**	**6.2**	**0.01**	**0.61**
T.Swan	**1441**	**2.7**	**0.72**	**1**	**4.8**	**0.04**	**-0.2**
T.Haines	**1440**	**1.9**	**0.96**	**1**	**5**	**0.98**	**1.14**
T.Agnew	**1441**	**1.7**	**0.59**	**1**	**4**	**-0.82**	**0.44**
T.Thurman	**1437**	**2**	**0.62**	**1**	**3**	**-0.42**	**0.08**
F.MD.Comp	**1076**	**0.001**	**0.92**	**-0.905**	**3.874**	**2.31**	**1.49**
F.Neut.Ccomp	**1077**	**-0.001**	**0.635**	**-0.985**	**2.489**	**0.81**	**0.64**
F.MNeut.Comp	**1076**	**0**	**0.687**	**-0.953**	**2.752**	**1.66**	**1.23**
F.Moore	**1079**	**1.814**	**1.179**	**1**	**7**	**4.52**	**2.08**
F.Swan	**1078**	**2.002**	**0.895**	**1**	**5**	**0.42**	**0.93**
F.Haines	**1078**	**3.584**	**1.627**	**1**	**5**	**-1.31**	**-0.62**
F.Agnew	**1079**	**1.296**	**0.498**	**1**	**3**	**2.30**	**1.77**
F.Thurman	**1080**	**1.52**	**0.674**	**1**	**4**	**1.45**	**1.4**

CWB	**1436**	**1.848**	**0.751**	**1**	**6.263**	**2.14**	**1.27**
CWBI	**1436**	**1.641**	**0.82**	**1**	**6.143**	**3.34**	**1.73**
CWBO	**1436**	**1.918**	**0.9**	**1**	**6.333**	**0.88**	**1.06**
Academic Dishonesty	**1436**	**1.345**	**0.472**	**1**	**4**	**4.63**	**2.03**
Infidelity Intentions	**668**	**2.447**	**1.072**	**1**	**6.143**	**0.72**	**0.98**
Attitudes	**1078**	**3.447**	**0.569**	**1.654**	**4.904**	**-0.21**	**-0.15**
Admissions	**1078**	**1.592**	**0.622**	**1**	**4.727**	**2.82**	**1.61**
Personality Integrity	**610**	**40.634**	**22.022**	**0**	**96**	**-0.60**	**0.22**
F.Attitudes.	**1079**	**4.209**	**0.648**	**1.25**	**4.923**	**1.72**	**-1.44**
F.Admissions	**1079**	**1.253**	**0.588**	**1**	**5**	**16.21**	**3.82**
F.Personality Integrity	**593**	**47.612**	**25.908**	**0**	**97**	**-0.97**	**-0.12**
BIDR	**360**	**4.267**	**0.644**	**2.075**	**6.5**	**0.56**	**0.32**
Conscientiousness	**1433**	**3.597**	**0.63**	**1.05**	**5**	**-0.12**	**-0.12**
Agreeableness	**1436**	**3.936**	**0.57**	**1.4**	**5**	**-0.09**	**-0.39**

Neuroticism	**1434**	**2.556**	**0.787**	**1**	**5**	**-0.31**	**0.16**
Openness \ Intellect	**1433**	**3.674**	**0.578**	**1.923**	**5**	**-0.30**	**-0.04**
Extraversion.	**1433**	**3.19**	**0.727**	**1**	**5**	**0.05**	**-0.28**
Orderliness	**1434**	**3.61**	**0.689**	**1**	**5**	**0.08**	**-0.29**
Industriousness	**1435**	**3.585**	**0.763**	**1.1**	**5**	**-0.32**	**-0.23**
Politeness	**1433**	**4.013**	**0.574**	**1.8**	**5**	**-0.12**	**-0.49**
Compassion	**1432**	**3.86**	**0.742**	**1**	**5**	**0.34**	**-0.63**
Withdrawal	**1433**	**2.715**	**0.902**	**1**	**5**	**-0.50**	**0.12**
Volatility	**1433**	**2.396**	**0.808**	**1**	**5**	**-0.10**	**0.42**
Open Facet	**1434**	**3.724**	**0.691**	**1.1**	**5**	**0.04**	**-0.33**
Intellect	**1433**	**3.625**	**0.696**	**1.4**	**5**	**-0.32**	**-0.18**
Assertiveness	**1435**	**3.071**	**0.845**	**1**	**5**	**-0.28**	**-0.3**
Enthusiasm	**1432**	**3.308**	**0.816**	**1**	**5**	**-0.22**	**-0.16**
ICAR (Cog Abil)	**1048**	**7.971**	**3.428**	**0**	**16**	**-0.68**	**0.03**
Psychopathy	**1437**	**2.025**	**0.638**	**1**	**4.444**	**-0.08**	**0.48**

Narcissism	**1437**	**2.539**	**0.694**	**1**	**5**	**-0.08**	**0.19**
Machiavellianism	**1437**	**2.96**	**0.754**	**1**	**5**	**-0.16**	**-0.16**
Impulsivity	**1437**	**1.834**	**0.499**	**1**	**3.846**	**0.37**	**0.62**
Age	**1432**	**35.156**	**13.022**	**18**	**84**	**-0.12**	**0.69**

MTurk

Variable	**N**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Kurtosis**	**Skew**
g	**367**	**-0.105**	**0.87**	**-1.936**	**2.163**	**3.18**	**0.68**
S_1	**367**	**0.089**	**0.898**	**-2.081**	**3.394**	**3.33**	**0.7**
S_2	**367**	**0.092**	**0.892**	**-1.912**	**3.187**	**3.36**	**0.24**
S_3	**367**	**-0.166**	**0.866**	**-1.864**	**4.545**	**10.7**	**1.91**
S_4	**367**	**0.084**	**0.679**	**-2.08**	**3.071**	**3.95**	**0.31**
S_5	**367**	**0.019**	**0.813**	**-2.361**	**2.616**	**3.19**	**0.37**
MD_Comp	**375**	**-0.085**	**0.921**	**-1.871**	**2.786**	**2.91**	**0.32**
Neut_Comp	**375**	**-0.074**	**0.751**	**-1.274**	**2.189**	**2.52**	**0.34**

MNeut_Comp	**375**	**-0.078**	**0.751**	**-1.513**	**2.19**	**2.82**	**0.31**
T.Moore	**375**	**2.3**	**1.1**	**1**	**6**	**3.69**	**0.96**
T.Swan	**375**	**2.7**	**0.77**	**1**	**4.8**	**2.8**	**-0.1**
T.Haines	**375**	**1.7**	**0.89**	**1**	**5**	**4.69**	**1.4**
T.Agnew	**375**	**1.7**	**0.55**	**1**	**3**	**2.16**	**0.38**
T.Thurman	**375**	**2**	**0.59**	**1**	**4**	**2.94**	**0.1**
F.MD.Comp	**187**	**0.085**	**1.023**	**-0.905**	**2.957**	**3.08**	**1.05**
F.Neut.Comp	**187**	**-0.301**	**0.833**	**-0.985**	**2.288**	**3.27**	**1.14**
F.MNeut.Comp	**187**	**-0.147**	**0.882**	**-0.953**	**2.368**	**3.27**	**1.14**
F.Moore	**187**	**1.949**	**1.282**	**1**	**6.375**	**4.18**	**1.45**
F.Swan	**187**	**2.052**	**0.97**	**1**	**5**	**2.76**	**0.79**
F.Haines	**187**	**1.775**	**1.176**	**1**	**5**	**3.81**	**1.42**
F.Agnew	**187**	**1.344**	**0.528**	**1**	**3**	**4.11**	**1.45**
F.Thurman	**187**	**1.596**	**0.739**	**1**	**3.714**	**2.86**	**1.03**
CWB	**375**	**1.935**	**0.829**	**1**	**6.263**	**6.01**	**1.41**

CWBI	**375**	**1.91**	**0.948**	**1**	**6.143**	**4.54**	**1.26**
CWBO	**375**	**1.839**	**0.918**	**1**	**6.333**	**4.94**	**1.32**
Academic Dishonesty	**375**	**1.315**	**0.489**	**1**	**3.583**	**8.51**	**2.33**
Attitudes	**187**	**3.526**	**0.521**	**1.808**	**4.558**	**2.77**	**-0.13**
Admissions	**187**	**1.534**	**0.601**	**1**	**3.909**	**5.11**	**1.57**
F..Attitudes.	**187**	**4.118**	**0.711**	**2.135**	**4.923**	**2.62**	**-0.91**
F.Admissions	**187**	**1.319**	**0.679**	**1**	**4.818**	**10.22**	**2.71**
BIDR	**188**	**4.315**	**0.755**	**2.075**	**6.5**	**3.09**	**0.2**
Conscientiousness	**374**	**3.677**	**0.638**	**1.05**	**5**	**3.31**	**-0.44**
Agreeableness	**374**	**3.891**	**0.629**	**1.4**	**5**	**3.17**	**-0.42**
Neuroticism	**373**	**2.433**	**0.823**	**1**	**4.95**	**2.7**	**0.39**
Openness \ Intellect	**373**	**3.639**	**0.608**	**1.923**	**5**	**2.65**	**-0.24**
Extraversion.	**374**	**3.096**	**0.791**	**1**	**4.95**	**2.75**	**-0.25**
Orderliness	**374**	**3.591**	**0.741**	**1**	**5**	**3.42**	**-0.57**
Industriousness	**374**	**3.762**	**0.722**	**1.1**	**5**	**2.97**	**-0.44**

Politeness	**374**	**4.006**	**0.643**	**1.8**	**5**	**2.75**	**-0.57**
Compassion	**374**	**3.776**	**0.807**	**1**	**5**	**3.55**	**-0.69**
Withdrawal	**373**	**2.563**	**0.941**	**1**	**5**	**2.49**	**0.39**
Volatility	**373**	**2.303**	**0.839**	**1**	**5**	**3.01**	**0.59**
Open Facet	**373**	**3.636**	**0.746**	**1.1**	**5**	**3.2**	**-0.47**
Intellect	**374**	**3.644**	**0.679**	**1.6**	**5**	**2.71**	**-0.34**
Assertiveness	**374**	**2.96**	**0.92**	**1**	**5**	**2.31**	**-0.28**
Enthusiasm	**374**	**3.233**	**0.872**	**1**	**5**	**2.71**	**-0.05**
ICAR (Cog Abil)	**375**	**8.008**	**3.093**	**1**	**14**	**2.23**	**-0.02**
Psychopathy	**375**	**2.022**	**0.698**	**1**	**4.444**	**2.79**	**0.54**
Narcissism	**375**	**2.464**	**0.773**	**1**	**5**	**2.62**	**0.32**
Machiavellianism	**375**	**3.005**	**0.825**	**1**	**4.889**	**2.57**	**-0.21**
Impulsivity	**375**	**1.728**	**0.451**	**1**	**3.615**	**3.31**	**0.57**
Age	**375**	**40.176**	**10.482**	**23**	**74**	**2.96**	**0.81**

REP

Variable	N	Mean	Std. Dev.	Min	Max	Kurtosis	Skew
g	322	0.086	0.783	-1.885	2.605	2.8	0.08
S_1	322	0.044	0.773	-1.723	3.46	3.6	0.68
S_2	322	-0.262	0.682	-2.061	2.447	3.39	-0.31
S_3	322	0.139	0.803	-1.717	3.511	4.95	0.85
S_4	322	-0.053	0.7	-2.097	2.086	3.4	-0.22
S_5	322	-0.038	0.848	-1.797	2.474	2.4	0.27
MD_Comp	328	-0.097	0.7	-1.871	1.604	2.85	-0.18
Neut_Comp	328	0.075	0.69	-1.274	1.904	2.53	0.11
MNeut_Comp	328	0.006	0.616	-1.513	1.548	2.6	-0.16
T.Moore	328	2.4	0.85	1	4.9	2.57	0.38
T.Swan	328	2.6	0.57	1	4.1	3.52	-0.62
T.Haines	328	2.1	0.84	1	5	3.29	0.61
T.Agnew	328	1.7	0.54	1	3	2.17	0.3
T.Thurman	328	2.1	0.54	1	3.7	2.72	-0.21

F.MD.Comp	**153**	**-0.035**	**0.725**	**-0.905**	**1.952**	**2.64**	**0.7**
F.Neut.Ccomp	**154**	**-0.357**	**0.703**	**-0.985**	**2.489**	**4.84**	**1.29**
F.MNeut.Comp	**153**	**-0.226**	**0.654**	**-0.953**	**2.245**	**3.68**	**1**
F.Moore	**154**	**1.75**	**0.9**	**1**	**5**	**4.31**	**1.41**
F.Swan	**155**	**1.989**	**0.784**	**1**	**3.769**	**1.92**	**0.3**
F.Haines	**154**	**1.634**	**0.894**	**1**	**5**	**4.77**	**1.49**
F.Agnew	**156**	**1.331**	**0.471**	**1**	**3**	**3.93**	**1.3**
F.Thurman	**156**	**1.567**	**0.642**	**1**	**4**	**4**	**1.14**
CWB	**328**	**1.677**	**0.585**	**1**	**4.158**	**4.91**	**1.3**
CWBI	**328**	**1.582**	**0.652**	**1**	**4.083**	**4.14**	**1.26**
CWBO	**328**	**1.635**	**0.774**	**1**	**4.833**	**4.65**	**1.45**
Academic Dishonesty	**328**	**1.367**	**0.427**	**1**	**3.833**	**7.49**	**1.8**
Attitudes	**156**	**3.442**	**0.467**	**2.212**	**4.596**	**2.73**	**0.1**
Admissions	**156**	**1.517**	**0.445**	**1**	**3.273**	**5.4**	**1.34**
F.Attitudes.	**156**	**4.09**	**0.591**	**2.231**	**4.923**	**2.88**	**-0.84**

F.Admissions	**156**	**1.232**	**0.416**	**0.909**	**4**	**22.68**	**3.91**
BIDR	**172**	**4.215**	**0.49**	**3.075**	**5.9**	**2.9**	**0.26**
Conscientiousness	**326**	**3.359**	**0.502**	**1.35**	**4.7**	**3.36**	**0**
Agreeableness	**328**	**3.96**	**0.477**	**2.3**	**5**	**2.94**	**-0.42**
Neuroticism	**327**	**2.821**	**0.571**	**1.1**	**4.8**	**4.01**	**0.37**
Openness \ Intellect	**326**	**3.529**	**0.483**	**2.2**	**4.8**	**2.93**	**0.39**
Extraversion.	**327**	**3.39**	**0.53**	**1.85**	**4.65**	**2.81**	**-0.02**
Orderliness	**327**	**3.507**	**0.584**	**1.5**	**4.9**	**2.78**	**0.09**
Industriousness	**327**	**3.209**	**0.625**	**1.2**	**4.8**	**3.19**	**-0.13**
Politeness	**325**	**3.915**	**0.502**	**2.3**	**5**	**2.83**	**-0.32**
Compassion	**324**	**4.009**	**0.582**	**1.4**	**5**	**3.48**	**-0.58**
Withdrawal	**326**	**2.986**	**0.643**	**1.1**	**4.8**	**3.38**	**0.07**
Volatility	**326**	**2.655**	**0.663**	**1**	**5**	**3.84**	**0.48**
Open Facet	**327**	**3.672**	**0.628**	**1.7**	**5**	**2.8**	**0.01**
Intellect	**325**	**3.386**	**0.536**	**2.1**	**4.8**	**2.79**	**0.14**

Assertiveness	**327**	**3.254**	**0.624**	**1.5**	**4.9**	**2.79**	**-0.09**
Enthusiasm	**326**	**3.525**	**0.632**	**1.4**	**5**	**3.19**	**-0.19**
ICAR (Cog Abil)	**328**	**8.14**	**3.318**	**0**	**16**	**2.45**	**-0.01**
Psychopathy	**328**	**2.019**	**0.528**	**1**	**4.333**	**3.49**	**0.41**
Narcissism	**328**	**2.7**	**0.472**	**1.556**	**4.222**	**3.08**	**0**
Machiavellianism	**328**	**2.892**	**0.604**	**1**	**4.889**	**3.08**	**-0.11**
Impulsivity	**328**	**2.024**	**0.481**	**1**	**3.846**	**3.78**	**0.65**
Age	**320**	**20.469**	**3.416**	**18**	**55**	**43.76**	**5.38**

Prolific

Variable	N	Mean	Std. Dev.	Min	Max	Kurtosis	Skew
g	**715**	**0.003**	**1.004**	**-1.888**	**2.909**	**3**	**0.6**
S_1	**715**	**-0.071**	**0.795**	**-2.427**	**3.311**	**3.72**	**0.44**
S_2	**715**	**0.067**	**0.91**	**-2.627**	**3.115**	**3.01**	**-0.05**
S_3	**715**	**0.008**	**1.038**	**-1.985**	**4.512**	**8.95**	**2.07**

S_4	**715**	**-0.04**	**0.762**	**-2.679**	**2.369**	**2.94**	**0.21**
S_5	**715**	**0.01**	**0.876**	**-2.118**	**2.975**	**3.05**	**0.63**
MD_Comp	**738**	**0.087**	**0.937**	**-1.871**	**3.044**	**2.73**	**0.06**
Neut_Comp	**734**	**0.006**	**0.867**	**-1.274**	**2.566**	**2.51**	**0.43**
MNeut_Comp	**734**	**0.04**	**0.823**	**-1.513**	**2.485**	**2.6**	**0.23**
T.Moore	**738**	**2.6**	**1.1**	**1**	**6.2**	**2.73**	**0.45**
T.Swan	**738**	**2.7**	**0.75**	**1**	**4.8**	**2.91**	**-0.24**
T.Haines	**737**	**1.9**	**1**	**1**	**5**	**3.97**	**1.22**
T.Agnew	**738**	**1.7**	**0.63**	**1**	**3**	**2.11**	**0.49**
T.Thurman	**734**	**2**	**0.67**	**1**	**3.9**	**2.34**	**0.14**
F.MD.Comp	**736**	**-0.013**	**0.928**	**-0.905**	**3.874**	**6.14**	**1.67**
F.Neut.Ccomp	**736**	**0.15**	**0.49**	**-0.985**	**2.051**	**5.15**	**1.31**
F.MNeut.Comp	**736**	**0.085**	**0.618**	**-0.953**	**2.752**	**5.87**	**1.61**
F.Moore	**738**	**1.794**	**1.201**	**1**	**7**	**8.53**	**2.28**
F.Swan	**736**	**1.992**	**0.899**	**1**	**5**	**3.76**	**1.04**

F.Haines	**737**	**4.45**	**0.99**	**1**	**5**	**6.64**	**-2.08**
F.Agnew	**736**	**1.276**	**0.496**	**1**	**3**	**5.99**	**1.95**
F.Thurman	**737**	**1.492**	**0.662**	**1**	**4**	**5.17**	**1.57**
CWB	**733**	**1.88**	**0.765**	**1**	**5**	**3.94**	**1.03**
CWBI	**733**	**1.53**	**0.787**	**1**	**6**	**8.05**	**2.11**
CWBO	**733**	**2.084**	**0.908**	**1**	**5.417**	**3.36**	**0.82**
Academic Dishonesty	**733**	**1.351**	**0.482**	**1**	**4**	**7.21**	**1.96**
Infidelity Intentions	**668**	**2.447**	**1.072**	**1**	**6.143**	**3.73**	**0.98**
T.EII.Attitudes	**735**	**3.427**	**0.599**	**1.654**	**4.904**	**2.71**	**-0.15**
T.EII.Admissions	**735**	**1.622**	**0.657**	**1**	**4.727**	**5.56**	**1.57**
T.EII.Social Desirable	**738**	**3.012**	**0.495**	**1.778**	**5**	**3.24**	**0.4**
Personality Integrity	**610**	**40.634**	**22.022**	**0**	**96**	**2.41**	**0.22**
F.EII.Attitudes.	**736**	**4.258**	**0.637**	**1.25**	**4.923**	**6.06**	**-1.75**
F.EII Admissions	**736**	**1.241**	**0.594**	**1**	**5**	**21.51**	**4.1**

F.EII.Social Desirable	**736**	**4.004**	**0.706**	**1**	**5**	**4.1**	**-0.98**
F.Personality Integrity	**593**	**47.612**	**25.908**	**0**	**97**	**2.04**	**-0.12**
Conscientiousness	**733**	**3.663**	**0.651**	**1.55**	**5**	**2.71**	**-0.18**
Agreeableness	**734**	**3.949**	**0.576**	**1.95**	**5**	**2.49**	**-0.31**
Neuroticism	**734**	**2.501**	**0.824**	**1**	**5**	**2.48**	**0.23**
Openness \ Intellect	**734**	**3.757**	**0.587**	**2**	**5**	**2.74**	**-0.14**
Extraversion.	**732**	**3.148**	**0.752**	**1**	**5**	**2.93**	**-0.16**
Orderliness	**733**	**3.665**	**0.701**	**1.2**	**5**	**2.87**	**-0.28**
Industriousness	**734**	**3.662**	**0.785**	**1.2**	**5**	**2.69**	**-0.36**
Politeness	**734**	**4.061**	**0.561**	**1.9**	**5**	**2.92**	**-0.52**
Compassion	**734**	**3.836**	**0.761**	**1.3**	**5**	**2.85**	**-0.48**
Withdrawal	**734**	**2.673**	**0.955**	**1**	**5**	**2.36**	**0.19**
Volatility	**734**	**2.328**	**0.827**	**1**	**5**	**2.76**	**0.47**
Open Facet	**734**	**3.793**	**0.682**	**1.3**	**5**	**2.88**	**-0.33**

Intellect	**734**	**3.72**	**0.742**	**1.4**	**5**	**2.75**	**-0.37**
Assertiveness	**734**	**3.047**	**0.877**	**1**	**5**	**2.63**	**-0.19**
Enthusiasm	**732**	**3.25**	**0.844**	**1**	**5**	**2.6**	**-0.08**
ICAR (Cog Abil)	**345**	**7.771**	**3.851**	**0**	**16**	**2.17**	**0.13**
Psychopathy	**734**	**2.029**	**0.652**	**1**	**4**	**2.72**	**0.44**
Narcissism	**734**	**2.505**	**0.724**	**1**	**4.778**	**2.89**	**0.32**
Machiavellianism	**734**	**2.968**	**0.776**	**1**	**5**	**2.82**	**-0.19**
Impulsivity	**734**	**1.803**	**0.508**	**1**	**3.692**	**3.24**	**0.66**
Age	**737**	**38.98**	**12.147**	**18**	**84**	**3.08**	**0.76**

All Group Means

	Prolific		**REP**		**MTurk**	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
g	**0.003**	**1.004**	**0.086**	**0.783**	**-0.105**	**0.87**
S_1	**-0.071**	**0.795**	**0.044**	**0.773**	**0.089**	**0.898**

S_2	**0.067**	**0.91**	**-0.262**	**0.682**	**0.092**	**0.892**
S_3	**0.008**	**1.038**	**0.139**	**0.803**	**-0.166**	**0.866**
S_4	**-0.04**	**0.762**	**-0.053**	**0.7**	**0.084**	**0.679**
S_5	**0.01**	**0.876**	**-0.038**	**0.848**	**0.019**	**0.813**
MD_Comp	**0.087**	**0.937**	**-0.097**	**0.7**	**-0.085**	**0.921**
Neut_Comp	**0.006**	**0.867**	**0.075**	**0.69**	**-0.074**	**0.751**
MNeut_Comp	**0.04**	**0.823**	**0.006**	**0.616**	**-0.078**	**0.751**
T.Moore	**2.6**	**1.1**	**2.4**	**0.85**	**2.3**	**1.1**
T.Swan	**2.7**	**0.75**	**2.6**	**0.57**	**2.7**	**0.77**
T.Haines	**1.9**	**1**	**2.1**	**0.84**	**1.7**	**0.89**
T.Agnew	**1.7**	**0.63**	**1.7**	**0.54**	**1.7**	**0.55**
T.Thurman	**2**	**0.67**	**2.1**	**0.54**	**2**	**0.59**
F.MD.Comp	**-0.013**	**0.928**	**-0.035**	**0.725**	**0.085**	**1.023**
F.Neut.Ccomp	**0.15**	**0.49**	**-0.357**	**0.703**	**-0.301**	**0.833**
F.MNeut.Comp	**0.085**	**0.618**	**-0.226**	**0.654**	**-0.147**	**0.882**

F.Moore	**1.794**	**1.201**	**1.75**	**0.9**	**1.949**	**1.282**
F.Swan	**1.992**	**0.899**	**1.989**	**0.784**	**2.052**	**0.97**
F.Haines	**4.45**	**0.99**	**1.634**	**0.894**	**1.775**	**1.176**
F.Agnew	**1.276**	**0.496**	**1.331**	**0.471**	**1.344**	**0.528**
F.Thurman	**1.492**	**0.662**	**1.567**	**0.642**	**1.596**	**0.739**
CWB	**1.88**	**0.765**	**1.677**	**0.585**	**1.935**	**0.829**
CWBI	**1.53**	**0.787**	**1.582**	**0.652**	**1.91**	**0.948**
CWBO	**2.084**	**0.908**	**1.635**	**0.774**	**1.839**	**0.918**
Academic Dishonesty	**1.351**	**0.482**	**1.367**	**0.427**	**1.315**	**0.489**
Attitudes	**3.427**	**0.599**	**3.442**	**0.467**	**3.526**	**0.521**
Admissions	**1.622**	**0.657**	**1.517**	**0.445**	**1.534**	**0.601**
F.Attitudes.	**4.258**	**0.637**	**4.09**	**0.591**	**4.118**	**0.711**
F.Admissions	**1.241**	**0.594**	**1.232**	**0.416**	**1.319**	**0.679**
BIDR	**NaN**		**4.215**	**0.49**	**4.315**	**0.755**
Conscientiousness	**3.663**	**0.651**	**3.359**	**0.502**	**3.677**	**0.638**

Agreeableness	**3.949**	**0.576**	**3.96**	**0.477**	**3.891**	**0.629**
Neuroticism	**2.501**	**0.824**	**2.821**	**0.571**	**2.433**	**0.823**
Openness \ Intellect	**3.757**	**0.587**	**3.529**	**0.483**	**3.639**	**0.608**
Extraversion.	**3.148**	**0.752**	**3.39**	**0.53**	**3.096**	**0.791**
Orderliness	**3.665**	**0.701**	**3.507**	**0.584**	**3.591**	**0.741**
Industriousness	**3.662**	**0.785**	**3.209**	**0.625**	**3.762**	**0.722**
Politeness	**4.061**	**0.561**	**3.915**	**0.502**	**4.006**	**0.643**
Compassion	**3.836**	**0.761**	**4.009**	**0.582**	**3.776**	**0.807**
Withdrawal	**2.673**	**0.955**	**2.986**	**0.643**	**2.563**	**0.941**
Volatility	**2.328**	**0.827**	**2.655**	**0.663**	**2.303**	**0.839**
Open Facet	**3.793**	**0.682**	**3.672**	**0.628**	**3.636**	**0.746**
Intellect	**3.72**	**0.742**	**3.386**	**0.536**	**3.644**	**0.679**
Assertiveness	**3.047**	**0.877**	**3.254**	**0.624**	**2.96**	**0.92**
Enthusiasm	**3.25**	**0.844**	**3.525**	**0.632**	**3.233**	**0.872**
ICAR (Cog Abil)	**7.771**	**3.851**	**8.14**	**3.318**	**8.008**	**3.093**

Psychopathy	**2.029**	**0.652**	**2.019**	**0.528**	**2.022**	**0.698**
Narcissism	**2.505**	**0.724**	**2.7**	**0.472**	**2.464**	**0.773**
Machiavellianism	**2.968**	**0.776**	**2.892**	**0.604**	**3.005**	**0.825**
Impulsivity	**1.803**	**0.508**	**2.024**	**0.481**	**1.728**	**0.451**
Age	**38.98**	**12.147**	**20.469**	**3.416**	**40.176**	**10.482**

Appendix 7

Subsample Regression Analyses

Prolific

	CWB		
	Null	Composite	Factor
ΔR^2	NA	0.033	0.033
Intercept	1.99**	1.94**	1.81**
Volatility	0	0.01	0
Withdrawal	0.02	0.02	0.03
Compassion	-0.03	-0.03	-0.03
Politeness	-0.19**	-0.15*	-0.13
Industriousness	-0.2**	-0.17**	-0.17**
Orderliness	0.03	0.05	0.04
Enthusiasm	0.01	0	0.01
Assertiveness	0.05	0.08	0.07
Intellect	0.03	0.02	0.03

Open_Facet	0.02	0.02	0.03
Machiavellianism	0.1*	0.03	0.04
Narcissism	-0.1	-0.12*	-0.11*
Psychopathy	0.25**	0.19**	0.18**
Impulsivity	0.13	0.12	0.14
Age	0	0.01**	0.01**
Gender	-0.04	-0.04	-0.06
MNeut	NA	0.22**	0.18**
N	713	712	693
R^2	0.26	0.293	0.293

MTurk

	CWB		
	Null	Composite	Factor
ΔR^2	NA	0.068	0.045

Intercept	1.13	0.85	0.97
Volatility	0.01	0.01	0.02
Withdrawal	-0.03	-0.02	-0.03
Compassion	0	-0.01	-0.01
Politeness	-0.08	0.03	0.01
Industriousness	-0.26**	-0.26**	-0.25**
Orderliness	0.15*	0.17**	0.16**
Enthusiasm	0.1	0.1	0.07
Assertiveness	-0.01	0.04	0.04
Intellect	0.05	0.08	0.1
Open_Facet	-0.07	-0.04	-0.05
Machiavellianism	0.11	0.01	0.04
Narcissism	-0.1	-0.14	-0.15*
Psychopathy	0.33**	0.21*	0.15
Impulsivity	0.45**	0.46**	0.47**

Age	0	0	0
Gender	-0.05	-0.02	-0.09
MNeut		0.43**	0.28**
N	371	371	363
R^2	0.341	0.409	0.386

REP

	CWB		
	Null	Composite	Factor
ΔR^2	NA	0.051	0.067
Intercept	-0.31	-0.18	-0.28
Volatility	-0.1	-0.11	-0.11
Withdrawal	0.13	0.12	0.12
Compassion	0.11	0.11	0.1
Politeness	-0.09	-0.07	-0.06

Industriousness	-0.05	-0.05	-0.05
Orderliness	0.03	0.04	0.04
Enthusiasm	0.05	0.05	0.04
Assertiveness	0.01	0.05	0.06
Intellect	0.07	0.04	0.04
Open_Facet	0.01	0	0.01
Machiavellianism	0.16*	0.13*	0.12
Narcissism	-0.04	-0.05	-0.04
Psychopathy	0.25**	0.17*	0.19*
Impulsivity	0.22**	0.19*	0.19*
Age	0.01	0.02**	0.02*
Gender	-0.09	-0.09	-0.1
MNeut	NA	0.24**	0.21**
N	300	300	295
R^2	0.235	0.286	0.302

Full Sample w ICAR

	CWB		
	Null	Composite	Factor
ΔR^2	NA	0.043	0.035
Intercept	1.1*	1.07*	1.1**
Volatility	-0.01	-0.01	-0.02
Withdrawal	0.05	0.06	0.06
Compassion	0	0	0
Politeness	-0.13*	-0.09	-0.09
Industriousness	-0.18**	-0.16**	-0.16**
Orderliness	0.07	0.09*	0.09*
Enthusiasm	0.08*	0.07*	0.07*
Assertiveness	0.05	0.08*	0.07
Intellect	-0.02	-0.02	0

Open_Facet	-0.02	-0.02	0
Machiavellianism	0.11**	0.05	0.06
Narcissism	-0.12**	-0.15**	-0.15**
Psychopathy	0.28**	0.2**	0.17**
Impulsivity	0.23**	0.22**	0.24**
ICAR	0	0	0
Age	0	0	0
Gender	-0.06	-0.05	-0.09*
MTurk	0.44**	0.41**	0.37**
Prolific	0.38**	0.31**	0.31**
MNeut	NA	0.27**	0.21**
N	1005	1005	982
R^2	0.288	0.331	0.323

Prolific

	CWBI		
	Null	Composite	Factor
ΔR^2	NA	0.008	0.002
Intercept	1.84**	1.81**	1.82**
Volatility	0.11*	0.11*	0.1
Withdrawal	-0.06	-0.06	-0.06
Compassion	0.02	0.02	0.01
Politeness	-0.32**	-0.29**	-0.28**
Industriousness	-0.02	-0.01	-0.01
Orderliness	0.03	0.04	0.04
Enthusiasm	0	-0.01	0.01
Assertiveness	0.11*	0.12*	0.1
Intellect	0.02	0.01	0.01
Open_Facet	-0.09	-0.09*	-0.08

Machiavellianism	0.07	0.03	0.04
Narcissism	-0.12*	-0.13*	-0.12*
Psychopathy	0.22**	0.19**	0.18**
Impulsivity	0.1	0.1	0.11
Age	0.01**	0.01**	0.01**
Gender	-0.04	-0.04	-0.06
MNeut	NA	0.11*	0.1**
N	713	712	693
R^2	0.212	0.22	0.214

MTurk

	CWBI		
	Null	Composite	Factor
ΔR^2	NA	0.051	0.019
Intercept	1.04	0.76	0.75

Volatility	-0.06	-0.06	-0.05
Withdrawal	0.02	0.03	0.01
Compassion	-0.02	-0.03	-0.03
Politeness	-0.05	0.06	0.05
Industriousness	-0.16	-0.16	-0.15
Orderliness	0.06	0.08	0.08
Enthusiasm	0.09	0.09	0.06
Assertiveness	0.07	0.12	0.13
Intellect	-0.11	-0.08	-0.06
Open_Facet	-0.04	-0.01	-0.01
Machiavellianism	0.15	0.05	0.1
Narcissism	-0.09	-0.13	-0.15
Psychopathy	0.37**	0.25*	0.21
Impulsivity	0.43**	0.43**	0.46**
Age	0	0	0

Gender	0	0.03	-0.02
MNeut		0.42**	0.25**
N	371	371	363
R^2	0.293	0.344	0.312

REP

	CWBI		
	Null	Composite	Factor
ΔR^2	NA	0.013	0.012
Intercept	0.22	0.3	0.24
Volatility	0.06	0.05	0.06
Withdrawal	-0.01	-0.01	-0.01
Compassion	0.06	0.06	0.05
Politeness	-0.07	-0.06	-0.05
Industriousness	0.02	0.01	0.03

Orderliness	0	0	-0.01
Enthusiasm	0.13	0.13	0.13
Assertiveness	-0.12	-0.09	-0.1
Intellect	-0.05	-0.07	-0.06
Open_Facet	0.15*	0.14*	0.14
Machiavellianism	0.06	0.04	0.04
Narcissism	-0.1	-0.1	-0.1
Psychopathy	0.34**	0.3**	0.31
Impulsivity	0.1	0.08	0.08
Age	0	0.01	0.01
Gender	-0.02	-0.02	-0.03
MNeut		0.14*	0.11
N	300	300	295
R^2	0.16	0.173	0.172

Full Sample w ICAR

	CWBI		
	Null	Composite	Factor
ΔR^2	NA	0.018	0.005
Intercept	0.92	0.91	0.96
Volatility	0.03	0.02	0.02
Withdrawal	0.01	0.01	0.01
Compassion	0.02	0.02	0.01
Politeness	-0.18**	-0.14*	-0.14*
Industriousness	-0.07	-0.06	-0.06
Orderliness	0.03	0.05	0.05
Enthusiasm	0.07	0.06	0.07
Assertiveness	0.07	0.09	0.07
Intellect	-0.1*	-0.1*	-0.09
Open_Facet	-0.02	-0.02	0

Machiavellianism	0.11**	0.05	0.07
Narcissism	-0.11*	-0.12*	-0.13*
Psychopathy	0.33**	0.27**	0.24**
Impulsivity	0.18*	0.17*	0.18**
ICAR	0.02*	0.02*	0.02*
Age	0	0	0
Gender	-0.02	-0.01	-0.04
MTurk	0.41**	0.39**	0.37**
Prolific	0.09	0.04	0.05
MNeut	NA	0.2**	0.14**
N	1005	1005	982
R^2	0.235	0.253	0.24

Prolific

	CWBO		
	Null	Composite	Factor
ΔR^2	NA	0.039	0.041
Intercept	2.08**	2.02*	1.8**
Volatility	-0.06	-0.05	-0.06
Withdrawal	0.07	0.07	0.08
Compassion	-0.05	-0.05	-0.05
Politeness	-0.11	-0.06	-0.03
Industriousness	-0.3**	-0.26**	-0.26**
Orderliness	0.02	0.05	0.04
Enthusiasm	0.02	0.01	0.01
Assertiveness	0.02	0.06	0.05
Intellect	0.04	0.03	0.04
Open_Facet	0.09	0.09	0.09
Machiavellianism	0.12*	0.03	0.04

Narcissism	-0.09	-0.11	-0.11
Psychopathy	0.26**	0.19**	0.19**
Impulsivity	0.15	0.14	0.17
Age	0	0	0
Gender	-0.04	-0.04	-0.06
MNeut	NA	0.29**	0.23**
N	713	712	693
R^2	0.256	0.295	0.297

MTurk

	CWBO		
	Null	Composite	Factor
ΔR^2	NA	0.049	0.036
Intercept	1.12	0.85	1.06

Volatility	0.08	0.08	0.1
Withdrawal	-0.06	-0.04	-0.04
Compassion	0.03	0.02	0.02
Politeness	-0.17	-0.07	-0.09
Industriousness	-0.26	-0.26*	-0.26*
Orderliness	0.22	0.24**	0.23**
Enthusiasm	0.12	0.12	0.1
Assertiveness	-0.07	-0.02	-0.03
Intellect	0.18	0.21*	0.25**
Open_Facet	-0.14	-0.12	-0.13
Machiavellianism	0.01	-0.08	-0.06
Narcissism	-0.09	-0.13	-0.13
Psychopathy	0.34	0.22*	0.13
Impulsivity	0.46	0.47**	0.47**
Age	0	0	0

Gender	-0.1	-0.07	-0.15
MNeut	NA	0.4**	0.29**
N	371	371	363
R^2	0.264	0.313	0.3

REP

	CWBO		
	Null	Composite	Factor
ΔR^2	NA	0.041	0.064
Intercept	-0.36	-0.21	-0.33
Volatility	-0.22*	-0.23**	-0.24**
Withdrawal	0.21*	0.2*	0.2*
Compassion	0.17	0.17	0.17
Politeness	-0.15	-0.12	-0.12

Industriousness	-0.07	-0.08	-0.09
Orderliness	0.06	0.08	0.09
Enthusiasm	-0.06	-0.06	-0.07
Assertiveness	0.15	0.2*	0.22*
Intellect	0.18	0.15	0.14
Open_Facet	-0.15	-0.17*	-0.15
Machiavellianism	0.21*	0.18*	0.16
Narcissism	-0.03	-0.03	-0.02
Psychopathy	0.18	0.08	0.11
Impulsivity	0.31**	0.28*	0.27*
Age	0.02	0.02*	0.03*
Gender	-0.22*	-0.21*	-0.24*
MNeut	NA	0.29**	0.26
N	300	300	295
R^2	0.196	0.237	0.26

Full Sample w ICAR

	CWBO		
	Null	Composite	Factor
ΔR^2	NA	0.038	0.035
Intercept	1.31*	1.29*	1.31*
Volatility	-0.03	-0.03	-0.04
Withdrawal	0.08	0.09	0.09
Compassion	0.01	0	0
Politeness	-0.15*	-0.1	-0.1
Industriousness	-0.23**	-0.21**	-0.2**
Orderliness	0.1*	0.12**	0.13**
Enthusiasm	0.09	0.08	0.07
Assertiveness	0.05	0.08	0.07
Intellect	0.05	0.06	0.08

Open_Facet	-0.06	-0.06	-0.05
Machiavellianism	0.09*	0.02	0.02
Narcissism	-0.14**	-0.16**	-0.17**
Psychopathy	0.25**	0.17**	0.12*
Impulsivity	0.26**	0.25**	0.27**
ICAR	-0.01	-0.01	-0.01
Age	0	0	0
Gender	-0.11*	-0.1	-0.15**
MTurk	0.43**	0.39**	0.34**
Prolific	0.62**	0.55**	0.55**
MNeut		0.3**	0.25**
N	1005	1005	982
R^2	0.243	0.281	0.278

Prolific

	AD		
	Null	Composite	Factor
ΔR^2	NA	0.059	0.058
Intercept	1.26**	1.22**	1.04**
Volatility	-0.01	-0.01	-0.01
Withdrawal	0.03	0.03	0.03
Compassion	0.04	0.04	0.03
Politeness	-0.09*	-0.05	-0.03
Industriousness	-0.07	-0.04	-0.04
Orderliness	0.03	0.05	0.05
Enthusiasm	0.06	0.05	0.05
Assertiveness	-0.01	0.02	0.02
Intellect	-0.06*	-0.07*	-0.06
Open_Facet	0.02	0.02	0.03

Machiavellianism	0.09**	0.03	0.04
Narcissism	-0.03	-0.05	-0.06
Psychopathy	0.19**	0.14**	0.12**
Impulsivity	-0.02	-0.03	-0.02
Age	0	0	0
Gender	-0.06	-0.06	-0.07
MNeut	NA	0.19**	0.17**
N	713	712	693
R^2	0.208	0.267	0.266

MTurk

	AD		
	Null	Composite	Factor
ΔR^2	NA	0.104	0.072
Intercept	0.85	0.65	0.72

Volatility	-0.04	-0.04	-0.04
Withdrawal	0.09	0.1*	0.08*
Compassion	0.01	0.01	0
Politeness	-0.11	-0.03	-0.03
Industriousness	0.03	0.03	0.02
Orderliness	0.04	0.05	0.05
Enthusiasm	0.1*	0.1**	0.08*
Assertiveness	-0.05	-0.01	0
Intellect	-0.07	-0.05	-0.04
Open_Facet	0	0.02	0.01
Machiavellianism	0.02	-0.06	-0.04
Narcissism	0.02	0	-0.02
Psychopathy	0.25**	0.15**	0.13*
Impulsivity	0.09	0.1	0.1
Age	0	0	0

Gender	-0.04	-0.02	-0.04
MNeut	NA	0.31**	0.22**
N	371	371	363
R^2	0.314	0.418	0.386

REP

	AD		
	Null	Composite	Factor
ΔR^2	NA	0.113	0.104
Intercept	0.72	0.87	0.76
Volatility	-0.04	-0.06	-0.06
Withdrawal	0.12*	0.11*	0.11*
Compassion	0.11	0.11*	0.11*
Politeness	-0.15	-0.12	-0.12
Industriousness	0.05	0.04	0.04

Orderliness	0.01	0.02	0.03
Enthusiasm	0	0	0
Assertiveness	-0.02	0.03	0.03
Intellect	0.01	-0.02	-0.03
Open_Facet	-0.04	-0.05	-0.04
Machiavellianism	0.06	0.03	0.02
Narcissism	0	-0.01	-0.01
Psychopathy	0.02	-0.07	-0.05
Impulsivity	0.26**	0.22**	0.22**
Age	-0.01**	-0.01	-0.01
Gender	0.07	0.08	0.06
MNeut	NA	0.27**	0.22**
N	300	300	295
R^2	0.16	0.273	0.264

Full Sample w ICAR

	AD		
	Null	Composite	Factor
ΔR^2	NA	0.09	0.089
Intercept	1.17**	1.15**	1.15**
Volatility	-0.03	-0.03	-0.03
Withdrawal	0.06*	0.06*	0.06*
Compassion	0.06*	0.06*	0.05*
Politeness	-0.14**	-0.1**	-0.09**
Industriousness	-0.02	-0.01	-0.01
Orderliness	0.03	0.05*	0.05*
Enthusiasm	0.07**	0.06**	0.05*
Assertiveness	-0.02	0.01	0.01
Intellect	-0.04	-0.04	-0.03
Open_Facet	-0.01	-0.01	0

Machiavellianism	0.08**	0.02	0.02
Narcissism	-0.01	-0.03	-0.05
Psychopathy	0.13**	0.06*	0.04
Impulsivity	0.09*	0.09*	0.09*
ICAR	-0.01**	-0.01**	-0.01**
Age	-0**	0	0
Gender	-0.03	-0.02	-0.03
MTurk	0.11*	0.09*	0.07
Prolific	0.13**	0.08	0.09*
MNeut	NA	0.24**	0.2**
N	1005	1005	982
R^2	0.23	0.32	0.319

Prolific

	Infidelity		
	Null	Composite	Factor
ΔR^2	NA	NA	NA
Intercept	0.25	0.21	0.19
Volatility	0.02	0.02	0
Withdrawal	0.09	0.08	0.08
Compassion	0.03	0.03	0.02
Politeness	-0.08	-0.05	-0.04
Industriousness	-0.05	-0.03	-0.03
Orderliness	0.04	0.06	0.05
Enthusiasm	0.13	0.12	0.14
Assertiveness	-0.08	-0.06	-0.07
Intellect	0.03	0.02	0.02
Open_Facet	-0.03	-0.03	0
Machiavellianism	0.14	0.09	0.1

Narcissism	0.09*	0.08	0.07
Psychopathy	0.58**	0.53**	0.5**
Impulsivity	-0.02	-0.02	-0.01
Age	0	0.01	0.01
Gender	0.05	0.04	0.03
MNeut	NA	0.17**	0.16**
N	650	649	632
R^2	0.203	0.213	213

Full Sample w ICAR

	Infidelity		
	Null	Composite	Factor
ΔR^2	NA	0.009	0.018
Intercept	0.84	0.78	0.83
Volatility	-0.06	-0.07	-0.09

Withdrawal	0.18	0.19	0.18
Compassion	-0.02	-0.03	-0.04
Politeness	-0.04	-0.01	-0.01
Industriousness	-0.15	-0.14	-0.15
Orderliness	0.15	0.17	0.18
Enthusiasm	0.12	0.11	0.14
Assertiveness	0.09	0.11	0.1
Intellect	-0.1	-0.08	-0.08
Open_Facet	-0.08	-0.08	-0.03
Machiavellianism	0.13	0.07	0.08
Narcissism	0.04	0.03	0.01
Psychopathy	0.68**	0.64**	0.61**
Impulsivity	-0.2	-0.19	-0.19
ICAR	0.01	0.01	0.01
Age	0	0	0

Gender	-0.04	-0.03	-0.07
MTurk	NA	NA	NA
Prolific	NA	NA	NA
MNeut	NA	0.16	0.18*
N	305	305	296
R^2	0.263	0.272	0.281

Appendix 8

Search Strings Used in Each Database Search

Business Source Premier
1. DE "NEUTRALIZATION theory"
2. (Employee) OR (Manager) OR (Organization) OR (Workplace)
3. 1 AND 2
4. Moral Disengagement
5. 2 AND 4
6. 3 OR 5

Sociological Abstracts
1. noft((Employee) OR (Manager) OR (Organization) OR (Workplace))
2. noft(neutralization)
3. 1 AND 2
4. noft(moral disengagement)
5. 1 AND 4
6. (1 AND 2) OR (1 AND 4)

Criminal Justice Database
1. noft(Neutralization)
2. noft((Employee) OR (Manager) OR (Organization) OR (Workplace))
3. 1 AND 2
4. noft(Moral disengagement)
5. 2 AND 4
6. (1 AND 2) OR (2 AND 4)

PsycINFO
1. Neutralization.mp.
2. (Employee or Manager or Organization or Workplace).mp.
3. 1 AND 2
4. Moral disengagement.mp.
5. 2 AND 4
6. 3 OR 5

Appendix 9

Search Strings to be Used in Update to Lee et al. (2020)

PsycINFO

1. Moral disengagement.mp
2. Academic dishonesty.mp
3. Cheating.mp
4. 2 or 3
5. 1 and 4

Results: 28

Academic Search Premier / Education Source / ERIC / OpenDissertations

1. TX Moral Disengagement
2. TX Academic Dishonesty
3. TX Academic Cheating
4. S2 OR S3
5. S1 AND S4

Results: 85

JSTOR

1. ((“Academic dishonesty”) AND (“moral disengagement”))

Results: 20

Google Scholar

1. "Moral disengagement" and "Academic Dishonesty"

Only selecting top 60 sorted by relevance

Appendix 10

Bandura et al. (1996) Measure of Moral Disengagement

Bandura et al., 1996 Moral Disengagement Items
It is alright to fight to protect your friends.
Slapping and shoving someone is just a way of joking.
Damaging some property is no big deal when you consider that others are beating people up.
A kid in a gang should not be blamed for the trouble the gang causes.
If kids are living under bad conditions they cannot be blamed for behaving aggressively.
It is okay to tell small lies because they don't really do any harm.
Some people deserve to be treated like animals.
If kids fight and misbehave in school it is their teacher's fault.
It is alright to beat someone who bad mouths your family.
To hit obnoxious classmates is just giving them "a lesson."
Stealing some money is not too serious compared to those who steal a lot of money.
A kid who only suggests breaking rules should not be blamed if other kids go ahead and do it.
If kids are not disciplined they should not be blamed for misbehaving.
Children do not mind being teased because it shows interest in them.
It is okay to treat badly somebody who behaved like a "worm."
If people are careless where they leave their things it is their own fault if they get stolen.
It is alright to fight when your group's honour is threatened.
Taking someone's bicycle without their permission is just "borrowing it."
It is okay to insult a classmate because beating him/her is worse.
If a group decides together to do something harmful it is unfair to blame any kid in the group for it.
Kids cannot be blamed for using bad words when all their friends do it.
Teasing someone does not really hurt them.
Someone who is obnoxious does not deserve to be treated like a human being.
Kids who get mistreated usually do things that deserve it.
It is alright to lie to keep your friends out of trouble.
It is not a bad thing to "get high" once in a while.
Compared to the illegal things people do, taking some things from a store without paying for them is not very serious.
It is unfair to blame a child who had only a small part in the harm caused by a group.

Kids cannot be blamed for misbehaving if their friends pressured them to do it.
Insults among children do not hurt anyone.
Some people have to be treated roughly because they lack feelings that can be hurt.
Children are not at fault for misbehaving if their parents force them too much.

Appendix 11

Detert et al. (2008) Measure of Moral Disengagement

Detert et al., 2008 Moral Disengagement Items
It is alright to fight to protect your friends.
It’s ok to steal to take care of your family’s needs.
It’s ok to attack someone who threatens your family's honor.
Sharing test questions is just a way of helping your friends.
Talking about people behind their backs is just part of the game.
Looking at a friend's homework without permission is just "barrowing it."
Damaging property is no big deal when you consider that others are beating up people.
Stealing some money is not too serious compared to those who steal a lot of money.
Compared to other illigal things that people do, taking some things from a store without paying for them is not very serious.
If people are living under bad conditions, they cannot be blamed for behaving aggresively.
If someone is pressured into doing something, they shouldn't be blamed for it.
People cannot be blamed for misbehaving if their friends pressured them to do it.
A member of a group of team should not be blamed for the trouble the team caused.
If a group decides together to do something harmful, it is unfair to blame any one member of the group for it.
You can't blame a person who plays only a small part in the harm caused by a group.
People don't mind being teased because it shows interest in them.
Teasing someone does not really hurt them.
Insults don't really hurt anyone.
If someone leaves something lying around, it's their own fault if it gets stolen.
People who are mistreated have usually done things to deserve it.
People are not at fault for misbehaving at work if their managers mistreat them.
Some people deserve to be treated like animals.
It is okay to treat badly someone who behaved like a "worm."
Someone who is obnoxious does not deserve to be treated like a human being.

Appendix 12

Neutralization and Moral Disengagement Scale and Item-Level Descriptives and Correlations

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max	Kurtosis	Skew
Moore_1	1440	2.3	1.5	1	1	3	7		
Moore_2	1441	2	1.3	1	1	2	7		
Moore_3	1438	2.9	1.7	1	1	4	7		
Moore_4	1441	2.9	1.7	1	1	4	7		
Moore_5	1441	2	1.3	1	1	2	7		
Moore_6	1441	1.8	1.2	1	1	2	7		
Moore_7	1441	2.1	1.5	1	1	3	7		
Moore_8	1440	2.4	1.5	1	1	3	7		
Swan_1	1441	3.3	1.2	1	2	4	5		
Swan_2	1441	2.6	1.2	1	2	4	5		
Swan_3	1440	3.2	1.2	1	2	4	5		
Swan_4	1439	3.4	1.2	1	2	4	5		
Swan_5	1440	2.2	0.99	1	1	3	5		
Swan_6	1440	3.4	1.2	1	2	4	5		
Swan_7	1438	2.1	0.97	1	1	2	5		
Swan_8	1439	2.7	1.2	1	2	4	5		
Swan_9	1440	2.7	1.1	1	2	4	5		
Swan_10	1439	2.9	1.2	1	2	4	5		

Swan_11	**1440**	**1.9**	**0.97**	**1**	**1**	**2**	**5**		
Swan_12	**1440**	**2**	**1.1**	**1**	**1**	**3**	**5**		
Swan_13	**1441**	**2.4**	**1.2**	**1**	**1**	**3**	**5**		
Thurman_1	**1436**	**1.7**	**0.72**	**1**	**1**	**2**	**4**		
Thurman_2	**1437**	**2.1**	**0.84**	**1**	**1**	**3**	**4**		
Thurman_3	**1435**	**2**	**0.78**	**1**	**1**	**3**	**4**		
Thurman_4	**1436**	**1.9**	**0.82**	**1**	**1**	**2**	**4**		
Thurman_5	**1436**	**2.4**	**0.86**	**1**	**2**	**3**	**4**		
Thurman_6	**1437**	**2.1**	**0.89**	**1**	**1**	**3**	**4**		
Thurman_7	**1436**	**2**	**0.84**	**1**	**1**	**3**	**4**		
Haines_1	**1440**	**1.8**	**1**	**1**	**1**	**2**	**5**		
Haines_2	**1438**	**2.1**	**1.2**	**1**	**1**	**3**	**5**		
Haines_3	**1440**	**1.8**	**1**	**1**	**1**	**2**	**5**		
Haines_4	**1440**	**2.2**	**1.2**	**1**	**1**	**3**	**5**		
Haines_5	**1438**	**2**	**1.2**	**1**	**1**	**2**	**5**		
Haines_6	**1438**	**1.9**	**1.1**	**1**	**1**	**2**	**5**		
Haines_7	**1435**	**2.2**	**1.2**	**1**	**1**	**3**	**5**		
Haines_8	**1435**	**1.7**	**0.97**	**1**	**1**	**2**	**5**		
Haines_9	**1436**	**1.8**	**1**	**1**	**1**	**2**	**5**		
Haines_10	**1440**	**1.8**	**1.1**	**1**	**1**	**2**	**5**		
Haines_11	**1440**	**1.9**	**1.1**	**1**	**1**	**2**	**5**		
Agnew_1	**1441**	**1.6**	**0.69**	**1**	**1**	**2**	**3**		

Agnew_2	**1440**	**1.6**	**0.7**	**1**	**1**	**2**	**3**		
Agnew_3	**1440**	**2**	**0.79**	**1**	**1**	**3**	**3**		
Agnew_4	**1441**	**1.8**	**0.74**	**1**	**1**	**2**	**3**		
Agnew_5	**1441**	**1.5**	**0.63**	**1**	**1**	**2**	**3**		
Moore_Full	**1441**	**2.5**	**1**	**1**	**1.6**	**3.2**	**6.2**	**0.01**	**0.61**
Swan_Full	**1441**	**2.7**	**0.72**	**1**	**2.2**	**3.2**	**4.8**	**0.04**	**-0.2**
Haines_Full	**1440**	**1.9**	**0.96**	**1**	**1**	**2.4**	**5**	**0.98**	**1.14**
Agnew_Full	**1441**	**1.7**	**0.59**	**1**	**1.2**	**2.2**	**3**	**-0.82**	**0.44**
Thurman_Full	**1437**	**2**	**0.62**	**1**	**1.6**	**2.4**	**4**	**-0.42**	**0.08**

Item and scale correlations can be viewed at the following link:

https://osf.io/va7nr/?view_only=038fd3c48e5a4ea8b1d79400646b7f22

Appendix 13

Link to the Meta-Analytic Dataset

https://osf.io/va7nr/?view_only=038fd3c48e5a4ea8b1d79400646b7f22

Appendix 14

Standardized Regression Coefficients for Regression Analyses Presented in Tables 17-22 and 26-28

	Study Three Incremental Validity Tables - Standardized Regression Coefficients		
	Predicting CWB Without Moral Neutralization		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.21	-0.43 – 0.01	0.06
Volatility Mean	-0.02	-0.09 – 0.05	0.561
Withdrawal Mean	0.05	-0.03 – 0.14	0.215
Compassion Mean	-0.01	-0.08 – 0.06	0.763
Politeness Mean	-0.1	-0.17 – -0.02	0.01
Industriousness Mean	-0.19	-0.27 – -0.10	<0.001
Orderliness Mean	0.05	-0.01 – 0.10	0.128
Enthusiasm Mean	0.05	-0.02 – 0.12	0.155
Assertiveness Mean	0.05	-0.03 – 0.12	0.235
Intellect Mean	0.03	-0.04 – 0.09	0.432
Open Facet Mean	-0.01	-0.06 – 0.05	0.751
Machiavellianism Mean	0.11	0.04 – 0.17	0.001
Narcissism Mean	-0.09	-0.16 – -0.02	0.01
Psychopathy Mean	0.24	0.17 – 0.31	<0.001
Impulsivity Mean	0.14	0.07 – 0.21	<0.001
Age	0.04	-0.02 – 0.10	0.173
Gender	-0.08	-0.18 – 0.02	0.132
Rep0MTurk1Prolific2 [1]	0.49	0.32 – 0.66	<0.001
Rep0MTurk1Prolific2 [2]	0.38	0.23 – 0.54	<0.001
	Predicting CWB With Moral Neutralization Composite		

Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.16	-0.37 – 0.06	0.153
Volatility Mean	-0.02	-0.09 – 0.05	0.581
Withdrawal Mean	0.05	-0.03 – 0.14	0.203
Compassion Mean	-0.01	-0.08 – 0.06	0.763
Politeness Mean	-0.06	-0.13 – 0.01	0.085
Industriousness Mean	-0.16	-0.24 – -0.08	<0.001
Orderliness Mean	0.06	0.01 – 0.12	0.03
Enthusiasm Mean	0.04	-0.03 – 0.10	0.239
Assertiveness Mean	0.08	0.01 – 0.15	0.03
Intellect Mean	0.02	-0.04 – 0.08	0.48
Open Facet Mean	-0.01	-0.06 – 0.04	0.713
Machiavellianism Mean	0.03	-0.03 – 0.10	0.274
Narcissism Mean	-0.11	-0.17 – -0.04	0.001
Psychopathy Mean	0.18	0.11 – 0.25	<0.001
Impulsivity Mean	0.14	0.07 – 0.20	<0.001
Age	0.08	0.02 – 0.14	0.007
Gender	-0.07	-0.17 – 0.03	0.144
Rep0MTurk1Prolific2 [1]	0.44	0.28 – 0.61	<0.001
Rep0MTurk1Prolific2 [2]	0.3	0.15 – 0.45	<0.001
MNeut comp	0.27	0.21 – 0.33	<0.001
	Predicting CWB With Moral Neutralization Factor		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.09	-0.31 – 0.12	0.386
Volatility Mean	-0.02	-0.09 – 0.05	0.545
Withdrawal Mean	0.05	-0.03 – 0.14	0.194
Compassion Mean	-0.01	-0.08 – 0.05	0.732
Politeness Mean	-0.05	-0.12 – 0.02	0.151

Industriousness Mean	-0.16	-0.24 – -0.07	<0.001
Orderliness Mean	0.06	0.00 – 0.12	0.039
Enthusiasm Mean	0.03	-0.03 – 0.10	0.294
Assertiveness Mean	0.07	-0.00 – 0.14	0.053
Intellect Mean	0.04	-0.03 – 0.10	0.26
Open Facet Mean	0	-0.06 – 0.05	0.935
Machiavellianism Mean	0.05	-0.01 – 0.11	0.118
Narcissism Mean	-0.11	-0.17 – -0.04	0.001
Psychopathy Mean	0.16	0.09 – 0.23	<0.001
Impulsivity Mean	0.15	0.08 – 0.21	<0.001
Age	0.08	0.03 – 0.14	0.005
Gender	-0.11	-0.21 – -0.01	0.027
Rep0MTurk1Prolific2 [1]	0.39	0.22 – 0.55	<0.001
Rep0MTurk1Prolific2 [2]	0.29	0.14 – 0.44	<0.001
g	0.26	0.20 – 0.32	<0.001
	Predicting CWBI Without Moral Neutralization		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0	-0.23 – 0.23	0.979
Volatility Mean	0.04	-0.03 – 0.12	0.278
Withdrawal Mean	-0.02	-0.11 – 0.07	0.695
Compassion Mean	0.01	-0.06 – 0.08	0.811
Politeness Mean	-0.12	-0.20 – -0.05	0.001
Industriousness Mean	-0.05	-0.14 – 0.03	0.233
Orderliness Mean	0.02	-0.04 – 0.08	0.5
Enthusiasm Mean	0.04	-0.03 – 0.11	0.243
Assertiveness Mean	0.08	-0.00 – 0.16	0.051
Intellect Mean	-0.03	-0.09 – 0.04	0.397
Open Facet Mean	-0.05	-0.10 – 0.01	0.121

Machiavellianism Mean	0.08	0.01 – 0.14	0.019
Narcissism Mean	-0.08	-0.15 – -0.01	0.022
Psychopathy Mean	0.24	0.16 – 0.31	<0.001
Impulsivity Mean	0.1	0.03 – 0.17	0.005
Age	0.09	0.02 – 0.15	0.008
Gender	-0.05	-0.16 – 0.05	0.331
Rep0MTurk1Prolific2 [1]	0.39	0.22 – 0.57	<0.001
Rep0MTurk1Prolific2 [2]	-0.04	-0.20 – 0.12	0.582
	Predicting CWBI With Moral Neutralization Composite		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.04	-0.19 – 0.26	0.756
Volatility Mean	0.04	-0.03 – 0.12	0.262
Withdrawal Mean	-0.02	-0.11 – 0.07	0.691
Compassion Mean	0.01	-0.06 – 0.08	0.809
Politeness Mean	-0.1	-0.18 – -0.03	0.007
Industriousness Mean	-0.04	-0.13 – 0.05	0.383
Orderliness Mean	0.03	-0.03 – 0.09	0.306
Enthusiasm Mean	0.04	-0.03 – 0.11	0.311
Assertiveness Mean	0.1	0.02 – 0.18	0.012
Intellect Mean	-0.03	-0.10 – 0.04	0.362
Open Facet Mean	-0.05	-0.10 – 0.01	0.112
Machiavellianism Mean	0.03	-0.03 – 0.10	0.326
Narcissism Mean	-0.09	-0.16 – -0.02	0.009
Psychopathy Mean	0.2	0.12 – 0.27	<0.001
Impulsivity Mean	0.1	0.03 – 0.17	0.006
Age	0.11	0.05 – 0.17	0.001
Gender	-0.05	-0.16 – 0.06	0.354

Rep0MTurk1Prolific2 [1]	0.36	0.19 – 0.54	<0.001
Rep0MTurk1Prolific2 [2]	-0.1	-0.26 – 0.06	0.225
MNeut comp	0.16	0.10 – 0.23	<0.001
	Predicting CWBI With Moral Neutralization Factor		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.09	-0.14 – 0.31	0.461
Volatility Mean	0.04	-0.03 – 0.12	0.259
Withdrawal Mean	-0.02	-0.11 – 0.06	0.597
Compassion Mean	0	-0.07 – 0.07	0.929
Politeness Mean	-0.1	-0.17 – -0.02	0.012
Industriousness Mean	-0.03	-0.12 – 0.05	0.447
Orderliness Mean	0.03	-0.03 – 0.09	0.356
Enthusiasm Mean	0.04	-0.03 – 0.11	0.273
Assertiveness Mean	0.09	0.01 – 0.16	0.026
Intellect Mean	-0.03	-0.09 – 0.04	0.451
Open Facet Mean	-0.03	-0.09 – 0.02	0.24
Machiavellianism Mean	0.05	-0.02 – 0.11	0.141
Narcissism Mean	-0.09	-0.16 – -0.03	0.008
Psychopathy Mean	0.18	0.10 – 0.25	<0.001
Impulsivity Mean	0.11	0.04 – 0.18	0.003
Age	0.11	0.05 – 0.17	0.001
Gender	-0.08	-0.18 – 0.03	0.147
Rep0MTurk1Prolific2 [1]	0.33	0.15 – 0.50	<0.001
Rep0MTurk1Prolific2 [2]	-0.1	-0.26 – 0.06	0.202
g	0.15	0.09 – 0.21	<0.001
	Predicting CWBO Without Moral Neutralization		
Predictors	*Estimates*	*CI*	*p*

(Intercept)	-0.3	-0.53 – -0.08	0.007
Volatility Mean	-0.05	-0.12 – 0.02	0.195
Withdrawal Mean	0.08	-0.00 – 0.17	0.052
Compassion Mean	-0.01	-0.08 – 0.06	0.754
Politeness Mean	-0.08	-0.16 – -0.01	0.028
Industriousness Mean	-0.2	-0.29 – -0.12	<0.001
Orderliness Mean	0.06	-0.00 – 0.12	0.059
Enthusiasm Mean	0.04	-0.03 – 0.11	0.24
Assertiveness Mean	0.02	-0.05 – 0.10	0.535
Intellect Mean	0.06	-0.00 – 0.12	0.063
Open Facet Mean	-0.01	-0.06 – 0.05	0.796
Machiavellianism Mean	0.08	0.02 – 0.14	0.009
Narcissism Mean	-0.08	-0.14 – -0.01	0.028
Psychopathy Mean	0.19	0.12 – 0.27	<0.001
Impulsivity Mean	0.13	0.06 – 0.20	<0.001
Age	0.01	-0.05 – 0.07	0.729
Gender	-0.09	-0.20 – 0.01	0.076
Rep0MTurk1Prolific2 [1]	0.42	0.25 – 0.59	<0.001
Rep0MTurk1Prolific2 [2]	0.64	0.48 – 0.79	<0.001
	Predicting CWBO With Moral Neutralization Composite		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.25	-0.47 – -0.03	0.024
Volatility Mean	-0.05	-0.12 – 0.02	0.199
Withdrawal Mean	0.08	0.00 – 0.17	0.046
Compassion Mean	-0.01	-0.08 – 0.06	0.757
Politeness Mean	-0.05	-0.12 – 0.02	0.171

Industriousness Mean	-0.18	-0.26 – -0.10	<0.001
Orderliness Mean	0.07	0.02 – 0.13	0.012
Enthusiasm Mean	0.03	-0.03 – 0.10	0.351
Assertiveness Mean	0.06	-0.02 – 0.13	0.121
Intellect Mean	0.06	-0.00 – 0.12	0.071
Open Facet Mean	-0.01	-0.06 – 0.05	0.76
Machiavellianism Mean	0.01	-0.05 – 0.07	0.695
Narcissism Mean	-0.09	-0.16 – -0.03	0.005
Psychopathy Mean	0.13	0.06 – 0.20	<0.001
Impulsivity Mean	0.13	0.06 – 0.19	<0.001
Age	0.05	-0.01 – 0.11	0.102
Gender	-0.09	-0.19 – 0.01	0.082
Rep0MTurk1Prolific2 [1]	0.37	0.21 – 0.54	<0.001
Rep0MTurk1Prolific2 [2]	0.55	0.40 – 0.70	<0.001
MNeut comp	0.26	0.20 – 0.32	<0.001
	Predicting CWBO With Moral Neutralization Factor		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.19	-0.41 – 0.03	0.085
Volatility Mean	-0.05	-0.12 – 0.02	0.172
Withdrawal Mean	0.09	0.01 – 0.17	0.035
Compassion Mean	-0.01	-0.07 – 0.06	0.8
Politeness Mean	-0.04	-0.11 – 0.03	0.274
Industriousness Mean	-0.18	-0.26 – -0.09	<0.001
Orderliness Mean	0.07	0.01 – 0.13	0.014
Enthusiasm Mean	0.02	-0.04 – 0.09	0.478
Assertiveness Mean	0.05	-0.02 – 0.13	0.164
Intellect Mean	0.07	0.01 – 0.13	0.023
Open Facet Mean	0	-0.06 – 0.05	0.872

Machiavellianism Mean	0.02	-0.04 – 0.08	0.51
Narcissism Mean	-0.1	-0.16 – -0.03	0.005
Psychopathy Mean	0.12	0.05 – 0.19	0.001
Impulsivity Mean	0.14	0.07 – 0.20	<0.001
Age	0.06	-0.00 – 0.12	0.065
Gender	-0.13	-0.23 – -0.02	0.015
Rep0MTurk1Prolific2 [1]	0.32	0.15 – 0.48	<0.001
Rep0MTurk1Prolific2 [2]	0.55	0.40 – 0.70	<0.001
g	0.26	0.20 – 0.32	<0.001
	Predicting Academic Dishonesty Without Moral Neutralization		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.06	-0.29 – 0.17	0.611
Volatility Mean	-0.05	-0.12 – 0.03	0.239
Withdrawal Mean	0.11	0.02 – 0.20	0.013
Compassion Mean	0.06	-0.01 – 0.13	0.118
Politeness Mean	-0.12	-0.20 – -0.05	0.002
Industriousness Mean	-0.05	-0.14 – 0.04	0.296
Orderliness Mean	0.05	-0.01 – 0.11	0.133
Enthusiasm Mean	0.11	0.04 – 0.18	0.003
Assertiveness Mean	-0.02	-0.10 – 0.06	0.578
Intellect Mean	-0.08	-0.14 – -0.01	0.026
Open Facet Mean	0.01	-0.05 – 0.06	0.847
Machiavellianism Mean	0.11	0.05 – 0.17	0.001
Narcissism Mean	-0.01	-0.08 – 0.06	0.727
Psychopathy Mean	0.23	0.16 – 0.31	<0.001
Impulsivity Mean	0.06	-0.01 – 0.14	0.073

Age	-0.08	-0.15 – -0.02	0.01
Gender	-0.07	-0.18 – 0.04	0.206
Rep0MTurk1Prolific2 [1]	0.17	-0.01 – 0.35	0.059
Rep0MTurk1Prolific2 [2]	0.24	0.07 – 0.40	0.004
	Predicting Academic Dishonesty With Moral Neutralization Composite		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.01	-0.21 – 0.23	0.906
Volatility Mean	-0.04	-0.11 – 0.03	0.235
Withdrawal Mean	0.11	0.03 – 0.20	0.009
Compassion Mean	0.06	-0.01 – 0.12	0.104
Politeness Mean	-0.08	-0.15 – -0.00	0.039
Industriousness Mean	-0.01	-0.10 – 0.07	0.75
Orderliness Mean	0.07	0.01 – 0.13	0.016
Enthusiasm Mean	0.09	0.03 – 0.16	0.006
Assertiveness Mean	0.03	-0.05 – 0.10	0.473
Intellect Mean	-0.08	-0.14 – -0.02	0.014
Open Facet Mean	0	-0.05 – 0.06	0.877
Machiavellianism Mean	0.01	-0.05 – 0.07	0.775
Narcissism Mean	-0.04	-0.11 – 0.03	0.257
Psychopathy Mean	0.14	0.07 – 0.21	<0.001
Impulsivity Mean	0.06	-0.01 – 0.12	0.1
Age	-0.03	-0.09 – 0.03	0.351
Gender	-0.06	-0.16 – 0.04	0.236
Rep0MTurk1Prolific2 [1]	0.11	-0.06 – 0.28	0.21
Rep0MTurk1Prolific2 [2]	0.11	-0.04 – 0.27	0.155
MNeut comp	0.38	0.32 – 0.44	<0.001

	Predicting Academic Dishonesty With Moral Neutralization Factor		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.04	-0.18 – 0.26	0.726
Volatility Mean	-0.04	-0.11 – 0.03	0.269
Withdrawal Mean	0.1	0.02 – 0.19	0.017
Compassion Mean	0.05	-0.02 – 0.12	0.142
Politeness Mean	-0.05	-0.13 – 0.02	0.167
Industriousness Mean	-0.01	-0.10 – 0.07	0.787
Orderliness Mean	0.07	0.01 – 0.13	0.023
Enthusiasm Mean	0.09	0.02 – 0.15	0.013
Assertiveness Mean	0.03	-0.04 – 0.11	0.39
Intellect Mean	-0.06	-0.13 – 0.00	0.052
Open Facet Mean	0.02	-0.04 – 0.07	0.571
Machiavellianism Mean	0.03	-0.04 – 0.09	0.42
Narcissism Mean	-0.05	-0.12 – 0.01	0.126
Psychopathy Mean	0.13	0.06 – 0.20	<0.001
Impulsivity Mean	0.07	-0.00 – 0.13	0.057
Age	-0.02	-0.08 – 0.04	0.445
Gender	-0.08	-0.19 – 0.02	0.115
Rep0MTurk1Prolific2 [1]	0.07	-0.10 – 0.24	0.415
Rep0MTurk1Prolific2 [2]	0.13	-0.03 – 0.28	0.107
g	0.38	0.32 – 0.44	<0.001
	Predicting Infidelity Intentions Without Moral Neutralization		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.06	-0.31 – 0.18	0.624
Volatility Mean	0.01	-0.10 – 0.13	0.797
Withdrawal Mean	0.07	-0.06 – 0.21	0.288
Compassion Mean	0.02	-0.08 – 0.12	0.655
Politeness Mean	-0.04	-0.15 – 0.07	0.47

Industriousness Mean	-0.04	-0.17 – 0.09	0.567
Orderliness Mean	0.03	-0.06 – 0.12	0.575
Enthusiasm Mean	0.1	-0.00 – 0.21	0.061
Assertiveness Mean	-0.07	-0.18 – 0.05	0.262
Intellect Mean	0.02	-0.08 – 0.11	0.711
Open Facet Mean	-0.02	-0.11 – 0.06	0.63
Machiavellianism Mean	0.1	0.01 – 0.19	0.034
Narcissism Mean	0.06	-0.04 – 0.16	0.259
Psychopathy Mean	0.34	0.24 – 0.45	<0.001
Impulsivity Mean	-0.01	-0.12 – 0.10	0.864
Age	0.05	-0.03 – 0.13	0.223
Gender	0.04	-0.12 – 0.20	0.604
	Predicting Infidelity Intentions With Moral Neutralization Composite		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.07	-0.32 – 0.17	0.565
Volatility Mean	0.02	-0.10 – 0.13	0.776
Withdrawal Mean	0.07	-0.07 – 0.20	0.318
Compassion Mean	0.02	-0.08 – 0.12	0.657
Politeness Mean	-0.03	-0.14 – 0.09	0.656
Industriousness Mean	-0.02	-0.16 – 0.11	0.712
Orderliness Mean	0.04	-0.05 – 0.13	0.422
Enthusiasm Mean	0.09	-0.01 – 0.20	0.079
Assertiveness Mean	-0.05	-0.16 – 0.07	0.415
Intellect Mean	0.01	-0.08 – 0.11	0.773
Open Facet Mean	-0.02	-0.11 – 0.06	0.627
Machiavellianism Mean	0.06	-0.04 – 0.16	0.218
Narcissism Mean	0.05	-0.05 – 0.15	0.324
Psychopathy Mean	0.32	0.21 – 0.42	<0.001
Impulsivity Mean	-0.01	-0.12 – 0.10	0.869
Age	0.07	-0.01 – 0.15	0.098
Gender	0.04	-0.12 – 0.20	0.617

MNeut comp	0.12	0.03 – 0.21	0.007
	Predicting Infidelity Intentions With Moral Neutralization Factor		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.06	-0.31 – 0.18	0.624
Volatility Mean	0.01	-0.10 – 0.13	0.797
Withdrawal Mean	0.07	-0.06 – 0.21	0.288
Compassion Mean	0.02	-0.08 – 0.12	0.655
Politeness Mean	-0.04	-0.15 – 0.07	0.47
Industriousness Mean	-0.04	-0.17 – 0.09	0.567
Orderliness Mean	0.03	-0.06 – 0.12	0.575
Enthusiasm Mean	0.1	-0.00 – 0.21	0.061
Assertiveness Mean	-0.07	-0.18 – 0.05	0.262
Intellect Mean	0.02	-0.08 – 0.11	0.711
Open Facet Mean	-0.02	-0.11 – 0.06	0.63
Machiavellianism Mean	0.1	0.01 – 0.19	0.034
Narcissism Mean	0.06	-0.04 – 0.16	0.259
Psychopathy Mean	0.34	0.24 – 0.45	<0.001
Impulsivity Mean	-0.01	-0.12 – 0.10	0.864
Age	0.05	-0.03 – 0.13	0.223
Gender	0.04	-0.12 – 0.20	0.604

	Predicting CWB without Integrity or Moral Neutralization		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.43	-0.58 – -0.27	<0.001
Conscientiousness	-0.2	-0.26 – -0.13	<0.001
Agreeableness	-0.25	-0.32 – -0.19	<0.001
Neuroticism	0.17	0.10 – 0.24	<0.001
Openness	0.03	-0.04 – 0.09	0.448
Extraversion	0.11	0.04 – 0.18	0.001

Source [1]	0.58	0.38 – 0.78	<0.001
Source [2]	0.48	0.31 – 0.64	<0.001
	Predicting CWB with True Overt Integrity Attitudes		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.38	-0.52 – -0.23	<0.001
Conscientiousness	-0.1	-0.16 – -0.03	0.003
Agreeableness	-0.11	-0.17 – -0.05	<0.001
Neuroticism	0.11	0.05 – 0.18	0.001
Openness	-0.03	-0.09 – 0.04	0.415
Extraversion	0.1	0.03 – 0.16	0.003
Source [1]	0.57	0.38 – 0.76	<0.001
Source [2]	0.41	0.25 – 0.56	<0.001
Overt true	-0.39	-0.45 – -0.33	<0.001
	Predicting CWB with Fake Overt Integrity Attitudes		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.42	-0.58 – -0.27	<0.001
Conscientiousness	-0.2	-0.26 – -0.13	<0.001
Agreeableness	-0.26	-0.33 – -0.20	<0.001
Neuroticism	0.17	0.11 – 0.24	<0.001
Openness	0.02	-0.05 – 0.09	0.545
Extraversion	0.12	0.05 – 0.19	0.001
Source [1]	0.58	0.38 – 0.79	<0.001
Source [2]	0.47	0.30 – 0.64	<0.001
Overt fake	0.03	-0.03 – 0.09	0.29
	Predicting CWB with True Overt Integrity Admissions		
Predictors	*Estimates*	*CI*	*p*

(Intercept)	-0.31	-0.45 – -0.17	<0.001
Conscientiousness	-0.12	-0.18 – -0.06	<0.001
Agreeableness	-0.15	-0.21 – -0.10	<0.001
Neuroticism	0.11	0.05 – 0.18	<0.001
Openness	0	-0.05 – 0.06	0.872
Extraversion	0.05	-0.01 – 0.12	0.09
Source [1]	0.49	0.31 – 0.68	<0.001
Source [2]	0.33	0.18 – 0.48	<0.001
Admissions true	0.43	0.37 – 0.48	<0.001
	Predicting CWB with Fake Overt Integrity Admissions		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.42	-0.57 – -0.26	<0.001
Conscientiousness	-0.19	-0.26 – -0.12	<0.001
Agreeableness	-0.25	-0.31 – -0.18	<0.001
Neuroticism	0.17	0.10 – 0.24	<0.001
Openness	0.03	-0.03 – 0.10	0.313
Extraversion	0.1	0.03 – 0.17	0.006
Source [1]	0.56	0.36 – 0.77	<0.001
Source [2]	0.47	0.30 – 0.63	<0.001
Admissions fake	0.07	0.01 – 0.12	0.016
	Predicting CWB with True Personality Based Integrity		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0	-0.08 – 0.08	1
Conscientiousness	-0.13	-0.23 – -0.03	0.008
Agreeableness	-0.17	-0.26 – -0.07	0.001
Neuroticism	0.13	0.03 – 0.22	0.012

Openness	0.05	-0.04 – 0.14	0.31
Extraversion	0	-0.10 – 0.10	0.982
Personality true	-0.24	-0.33 – -0.14	<0.001
	Predicting CWB with Fake Personality Based Integrity		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0	-0.08 – 0.08	1
Conscientiousness	-0.17	-0.27 – -0.07	0.001
Agreeableness	-0.28	-0.37 – -0.19	<0.001
Neuroticism	0.17	0.07 – 0.26	0.001
Openness	0.07	-0.02 – 0.17	0.124
Extraversion	0.09	-0.01 – 0.19	0.074
Personality fake	0.01	-0.07 – 0.09	0.823
	Predicting CWB with True Moral Neutralization		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.38	-0.52 – -0.23	<0.001
Conscientiousness	-0.13	-0.20 – -0.06	<0.001
Agreeableness	-0.13	-0.19 – -0.06	<0.001
Neuroticism	0.14	0.07 – 0.20	<0.001
Openness	0	-0.06 – 0.07	0.899
Extraversion	0.08	0.02 – 0.15	0.015
Source [1]	0.55	0.36 – 0.75	<0.001
Source [2]	0.41	0.25 – 0.57	<0.001
mneut true	0.31	0.25 – 0.37	<0.001
	Predicting CWB with Fake Moral Neutralization		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.41	-0.57 – -0.26	<0.001

Conscientiousness	-0.2	-0.26 – -0.13	<0.001
Agreeableness	-0.24	-0.31 – -0.18	<0.001
Neuroticism	0.17	0.10 – 0.24	<0.001
Openness	0.03	-0.04 – 0.09	0.387
Extraversion	0.11	0.04 – 0.18	0.003
Source [1]	0.57	0.37 – 0.78	<0.001
Source [2]	0.46	0.28 – 0.63	<0.001
mneut fake	0.03	-0.02 – 0.09	0.25
	Predicting Academic Dishonesty without Integrity or Moral Neutralization		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.11	-0.27 – 0.05	0.179
Conscientiousness	-0.1	-0.17 – -0.03	0.007
Agreeableness	-0.2	-0.26 – -0.13	<0.001
Neuroticism	0.18	0.10 – 0.25	<0.001
Openness	-0.04	-0.11 – 0.03	0.285
Extraversion	0.17	0.09 – 0.24	<0.001
Source [1]	0.12	-0.10 – 0.33	0.287
Source [2]	0.13	-0.05 – 0.31	0.146
	Predicting Academic Dishonesty with True Overt Integrity Attitudes		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.06	-0.21 – 0.09	0.455
Conscientiousness	0	-0.07 – 0.07	0.955
Agreeableness	-0.05	-0.11 – 0.02	0.154
Neuroticism	0.12	0.05 – 0.18	0.001
Openness	-0.09	-0.16 – -0.03	0.006
Extraversion	0.15	0.08 – 0.22	<0.001

Source [1]	0.11	-0.09 – 0.31	0.292
Source [2]	0.06	-0.11 – 0.22	0.506
Overt true	-0.41	-0.47 – -0.35	<0.001
	Predicting Academic Dishonesty with Fake Overt Integrity Attitudes		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.12	-0.28 – 0.04	0.137
Conscientiousness	-0.1	-0.17 – -0.03	0.007
Agreeableness	-0.16	-0.23 – -0.09	<0.001
Neuroticism	0.17	0.10 – 0.24	<0.001
Openness	-0.01	-0.08 – 0.05	0.669
Extraversion	0.13	0.05 – 0.20	0.001
Source [1]	0.11	-0.10 – 0.32	0.319
Source [2]	0.15	-0.03 – 0.32	0.097
Overt fake	-0.14	-0.21 – -0.08	<0.001
	Predicting Academic Dishonesty with True Overt Integrity Admissions		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.01	-0.14 – 0.15	0.92
Conscientiousness	-0.02	-0.09 – 0.04	0.486
Agreeableness	-0.1	-0.16 – -0.04	0.002
Neuroticism	0.12	0.05 – 0.18	0.001
Openness	-0.06	-0.12 – 0.00	0.068
Extraversion	0.11	0.04 – 0.17	0.002
Source [1]	0.03	-0.17 – 0.22	0.785
Source [2]	-0.02	-0.18 – 0.14	0.829
Admissions true	0.43	0.37 – 0.48	<0.001

	Predicting Academic Dishonesty with Fake Overt Integrity Admissions		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.08	-0.24 – 0.07	0.296
Conscientiousness	-0.09	-0.16 – -0.02	0.014
Agreeableness	-0.18	-0.25 – -0.12	<0.001
Neuroticism	0.17	0.10 – 0.24	<0.001
Openness	-0.02	-0.09 – 0.05	0.616
Extraversion	0.13	0.06 – 0.21	<0.001
Source [1]	0.07	-0.14 – 0.28	0.505
Source [2]	0.1	-0.07 – 0.28	0.239
Admissions fake	0.16	0.11 – 0.22	<0.001
	Predicting Academic Dishonesty with True Personality Based Integrity		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0	-0.08 – 0.08	1
Conscientiousness	-0.04	-0.14 – 0.06	0.47
Agreeableness	-0.07	-0.18 – 0.03	0.165
Neuroticism	0.17	0.07 – 0.27	0.001
Openness	-0.04	-0.14 – 0.05	0.388
Extraversion	0.02	-0.09 – 0.12	0.754
Personality true	-0.21	-0.31 – -0.11	<0.001
	Predicting Academic Dishonesty with Fake Personality Based Integrity		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0	-0.08 – 0.08	1
Conscientiousness	-0.06	-0.16 – 0.04	0.246
Agreeableness	-0.18	-0.27 – -0.08	<0.001
Neuroticism	0.2	0.10 – 0.31	<0.001
Openness	-0.02	-0.11 – 0.08	0.705

Extraversion	0.08	-0.03 – 0.18	0.153
Personality fake	-0.09	-0.17 – -0.01	0.028
	Predicting Academic Dishonesty with True Moral Neutralization		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.04	-0.19 – 0.11	0.61
Conscientiousness	0	-0.07 – 0.06	0.921
Agreeableness	-0.02	-0.09 – 0.04	0.527
Neuroticism	0.13	0.06 – 0.19	<0.001
Openness	-0.07	-0.13 – -0.00	0.035
Extraversion	0.12	0.06 – 0.19	<0.001
Source [1]	0.08	-0.12 – 0.27	0.455
Source [2]	0.04	-0.13 – 0.20	0.656
mneut true	0.44	0.38 – 0.50	<0.001
	Predicting Academic Dishonesty with Fake Moral Neutralization		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.04	-0.20 – 0.13	0.667
Conscientiousness	-0.1	-0.17 – -0.02	0.008
Agreeableness	-0.16	-0.22 – -0.09	<0.001
Neuroticism	0.17	0.10 – 0.24	<0.001
Openness	-0.02	-0.09 – 0.05	0.579
Extraversion	0.13	0.06 – 0.20	<0.001
Source [1]	0.08	-0.13 – 0.29	0.454
Source [2]	0.03	-0.15 – 0.21	0.734
mneut fake	0.18	0.12 – 0.24	<0.001
	Predicting Infidelity Intentions without Integrity or Moral Neutralization		
Predictors	*Estimates*	*CI*	*p*

(Intercept)	0	-0.07 – 0.07	1
Conscientiousness	-0.03	-0.13 – 0.06	0.461
Agreeableness	-0.23	-0.31 – -0.14	<0.001
Neuroticism	0.19	0.09 – 0.28	<0.001
Openness	0.03	-0.06 – 0.12	0.53
Extraversion	0.18	0.08 – 0.27	<0.001
	Predicting Infidelity Intentions with True Overt Integrity Attitudes		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0	-0.07 – 0.07	1
Conscientiousness	0.02	-0.07 – 0.12	0.604
Agreeableness	-0.14	-0.23 – -0.05	0.003
Neuroticism	0.13	0.04 – 0.23	0.006
Openness	-0.01	-0.09 – 0.08	0.882
Extraversion	0.16	0.07 – 0.26	0.001
Overt true	-0.25	-0.34 – -0.17	<0.001
	Predicting Infidelity Intentions with Fake Overt Integrity Attitudes		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0	-0.07 – 0.07	1
Conscientiousness	-0.04	-0.13 – 0.06	0.449
Agreeableness	-0.22	-0.31 – -0.13	<0.001
Neuroticism	0.18	0.09 – 0.28	<0.001
Openness	0.03	-0.06 – 0.12	0.47
Extraversion	0.17	0.07 – 0.27	0.001
Overt fake	-0.03	-0.11 – 0.05	0.444
	Predicting Infidelity Intentions with True Overt Integrity Admissions		
Predictors	*Estimates*	*CI*	*p*

(Intercept)	0	-0.07 – 0.07	1
Conscientiousness	0.01	-0.08 – 0.10	0.767
Agreeableness	-0.15	-0.23 – -0.06	0.001
Neuroticism	0.14	0.04 – 0.23	0.004
Openness	0.01	-0.07 – 0.10	0.772
Extraversion	0.13	0.04 – 0.22	0.004
Admissions true	0.3	0.22 – 0.37	<0.001
	Predicting Infidelity Intentions with Fake Overt Integrity Admissions		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0	-0.07 – 0.07	1
Conscientiousness	-0.03	-0.13 – 0.06	0.482
Agreeableness	-0.23	-0.31 – -0.14	<0.001
Neuroticism	0.19	0.09 – 0.28	<0.001
Openness	0.03	-0.06 – 0.12	0.496
Extraversion	0.17	0.08 – 0.27	<0.001
Admissions fake	0.04	-0.04 – 0.11	0.342
	Predicting Infidelity Intentions with True Personality Based Integrity		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0	-0.08 – 0.08	1
Conscientiousness	0.01	-0.09 – 0.12	0.798
Agreeableness	-0.09	-0.20 – 0.02	0.128
Neuroticism	0.17	0.06 – 0.28	0.003
Openness	0	-0.10 – 0.11	0.933
Extraversion	0.07	-0.04 – 0.19	0.211
Personality true	-0.23	-0.33 – -0.13	<0.001
	Predicting Infidelity Intentions with Fake Personality Based Integrity		

Predictors	Estimates	CI	p
(Intercept)	0	-0.09 – 0.09	1
Conscientiousness	-0.01	-0.12 – 0.09	0.817
Agreeableness	-0.2	-0.30 – -0.10	<0.001
Neuroticism	0.21	0.10 – 0.32	<0.001
Openness	0.03	-0.08 – 0.13	0.608
Extraversion	0.15	0.03 – 0.26	0.01
Personality fake	-0.07	-0.16 – 0.01	0.091
	Predicting Infidelity Intentions with True Moral Neutralization		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0	-0.07 – 0.07	1
Conscientiousness	0.02	-0.07 – 0.11	0.647
Agreeableness	-0.14	-0.22 – -0.05	0.003
Neuroticism	0.14	0.05 – 0.24	0.003
Openness	0.01	-0.07 – 0.10	0.784
Extraversion	0.16	0.06 – 0.25	0.001
mneut true	0.26	0.17 – 0.34	<0.001
	Predicting Infidelity Intentions with Fake Moral Neutralization		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0	-0.07 – 0.07	1
Conscientiousness	-0.04	-0.13 – 0.06	0.45
Agreeableness	-0.21	-0.30 – -0.13	<0.001
Neuroticism	0.18	0.09 – 0.28	<0.001
Openness	0.04	-0.05 – 0.13	0.396
Extraversion	0.16	0.07 – 0.26	0.001
mneut fake	0.11	0.04 – 0.19	0.002