Faking on Self- and Informant-reports of the Dark Triad of Personality

Sarah A. Walker

School of Psychology
Faculty of Science
The University of Sydney

0000-0002-6767-8604

A thesis submitted in fulfilment of the requirements for the degree of Master of Philosophy
(MPhil) 2021
Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work.  
This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all 
the assistance received in preparing this thesis and sources have been acknowledged.

Signature:

Name: Sarah A. Walker
Ethical Approval

All studies presented in this thesis were carried out in accordance with the National Statement on Ethical Conduct in Human Research (2007) and received ethical approval from The University of Sydney Human Research Ethics Committee(s) prior to their commencement under protocol 2019/375.
Author Attribution Statement

Chapter 2 presents data from a meta-analysis currently under peer-review as [Walker, S. A., Birney, D. P., & MacCann, C. (2018). *How much can people fake on the Dark Triad? A meta-analysis and systematic review of instructed faking.* (Under Review)]. I led the development of the review protocol, conducted the literature search, extracted data from the final selection of included studies, performed the statistical analyses, synthesised the results and wrote the manuscript under the guidance of Associate Professor Carolyn MacCann. C. MacCann also assisted in data extraction to allow for assessment of inter-rater reliability. Feedback on the final manuscript was provided by all co-authors.


As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Acknowledgements

“No individual is alone responsible for a single stepping stone along the path of progress, and where the path is smooth progress is most rapid” - Ernest Lawrence

Undertaking this master’s research has been an incredible experience which would not have been possible without the support and guidance I have received from many people.

First and foremost, I am extremely grateful to my supervisors, Associate Professors Carolyn MacCann, and Damian Birney for their generous advice, continuous support, and guidance throughout this process. Their passion and dedication to good science has been a continuous inspiration. Carolyn, your generosity, unwavering support, and expert guidance has not only deepened my love of research but has challenged me to continue learning and to go further than I thought I could. Damian, 5-years ago you took a chance on a first year undergraduate and I couldn’t be more grateful for that opportunity. Your passion for discovery, generosity, and kindness have been instrumental in developing my love of research that all started with you giving me a huge reading list.

I am indebted to my friends and colleagues in the School of Psychology, past and present, who have made this experience such a joy, who have challenged me, distracted me with long coffee dates, chats and laughter. Heartfelt thanks (in alphabetical order) to Hannah, Hayley, Indako, Joel, Jordan, Kate, Kelly, Kit, Natali, and Sarah D. Special thanks to Kit for showing me the ropes 5-years ago and for your friendship ever since. To Indako, whose friendship grew from where all should: pasta, wine, and passeggiate while conferencing in Florence, and whose honest appraisal of all things HDR related has helped me stay on track. To Kelly and Natali, your unwavering support, and friendship means more to me than I could ever express in words.

To my family and friends. I am extremely grateful to my parents, Alan and Barbara, who have been a constant source of encouragement. I couldn’t have wished for two finer people to teach me about life, love, sacrifice, and joy. To my in-laws, Geoff and Margaret, who consistently checked in on me.

Finally, to my darling Liam. Husband, best friend, fiercest supporter. Thank you for always finding a way to make me smile, for never letting me give up when the learning curve was steep, for staying up with me into the small hours when R wouldn’t do as it was told, for ensuring I always had coffee, and for always believing in me.
# Table of Contents

Abstract 1  
List of Tables 2  
List of Figures 3  
Presentation Format 4  

## CHAPTER 1: Literature Review 7  
1.1 Self-report Measures of Personality 8  
1.2 Response Distortion 9  
1.3 Informant-report Measures of Personality 14  
1.4 Evaluative and Descriptive Item Content 16  
1.5 Self-interest and Other-interest 18  
1.6 The Dark Triad 19  
1.7 Proposed Research on Measuring Faking on Self- and Informant-reports of the Dark Triad 24  
1.8 Overview and Aims of the Thesis 24  

2.1 Abstract 26  
2.2 Introduction 27  
2.3 Method 35  
2.4 Results 37  
2.4.1 Systematic Review Summary 38  
2.4.2 Meta-analysis 40  
2.5 Discussion 46  

## CHAPTER 3: Faking Good and Bad on Self-Reports versus Informant-Reports of Dark Triad Personality 67  
3.1 Abstract 67  
3.2 Introduction 68  
3.3 Study 1: Instructed Faking on Self- and Informant-reports 76  
3.3.1 Method 77  
3.3.2 Results 81
3.3.3 Discussion

3.4 Study 2: Expert Ratings of Evaluative Content in Dark Triad Measures
    3.4.1 Method
    3.4.2 Results
    3.4.3 Discussion

3.5 General Discussion

CHAPTER 4: General Discussion

4.1 Implications for Research and Theory
4.2 Limitations and Future Directions
4.3 Concluding Statements

REFERENCES

APPENDIX A: Instructed Faking Instructions
APPENDIX B: Manipulation Checks
APPENDIX C: Pre-registration for Chapter 3
APPENDIX D: Chapter 3 Data
Abstract

Self-report rating scales are the most commonly used method for obtaining information about an individual’s personality and are used widely. However, prior meta-analytic results demonstrate people can distort (or fake) their responses on Big Five personality scales when instructed to do so. Informant-reports (where a knowledgeable informant rates a target’s personality) are an alternate method and are often used to verify the accuracy of self-reports. As yet, it is unclear whether informants can also fake on personality tests. The purpose of this thesis was to extend prior findings on instructed faking in three ways: 1) examining both self- and informant reports, 2) extending research from the Big Five to the Dark Triad domains and facets, and 3) examining whether item-level social desirability accounts for score changes due to faking. Chapter 1 introduces background theory and research on response distortion and the Dark Triad. Chapter 2 presents a meta-analysis and systematic review ($k = 13$) summarising mean score changes on Dark Triad domains and facets when people are instructed to ‘fake good’ or ‘fake bad’. Due to a small $k$, narcissism and Machiavellianism were systematically reviewed. Results indicate substantial mean score changes between the “answer honestly” and “faking” conditions, with larger differences for ‘fake bad’ than ‘fake good’. Chapter 3 consists of two studies. Study 1 ($N = 834$) examines instructed faking on the Dark Triad in a 2x3 between-person design crossing target (self- versus informant-report) and instructions (answer honestly, fake good, or fake bad). Results demonstrate significant faking good and faking bad, with similar effect sizes for self- and informant-reports. Study 2 obtained expert ratings ($N = 9$) of social desirability of the Dark Triad items used in Study 1. Results indicated: a) Dark Triad scales were generally socially undesirable, but the desirability differed across facets; and b) the extent of faking could be largely accounted for by item social desirability at both the item and scale level. Implications for the measurement of the Dark Triad, and personality more broadly, are discussed in the closing Chapter (Chapter 4) alongside recommendations for future research.
## List of Tables

| Table 2.1 | Included Studies and Study Characteristics for the Systematic Review and Meta-Analysis of Instructed Faking on the Dark Triad of Personality | 42 |
| Table 2.2 | Meta-analytic standardized mean differences (fake - honest) for the effect of instructed faking on the domain and facets of psychopathy. | 44 |
| Table 3.1 | Reliability and descriptive statistics for each dark triad domain and facet under answer honestly, fake good, and fake bad instruction conditions for self-report and informant-report (Cohen’s d compares faking conditions to answer honestly) | 83 |
| Table 3.2 | Results of the 3x2 between-subjects ANOVA testing instruction condition by target (self-report versus informant-report) and interactions. | 87 |
| Table 3.3 | Standardized beta-weights for regressions predicting dark triad traits from contrast-coded conditions (self/informant, and fake-good/fake-bad/honest), adding self-interest, (plus self-interest interactions) or other interest (and other-interest interactions) | 89 |
| Table 3.4 | Reliability and descriptive statistics for social desirability ratings by scale. | 95 |
| Table 3.5 | Correlations of expert social desirability ratings and participant's mean change in faking for self-reports, and informant-reports. | 96 |
# List of Figures

<table>
<thead>
<tr>
<th>Figure 2.1</th>
<th>Quorum Chart of Studies Included in the Systematic Review and Meta-analysis</th>
<th>36</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.2</td>
<td>Fake Good and Fake Bad Funnel Plot</td>
<td>46</td>
</tr>
</tbody>
</table>
**Presentation Format**

This thesis presents a combination of submitted and not-presently submitted manuscripts. Each chapter is intended to be read both as a stand-alone work and as a continuation of the ideas expressed within this thesis.
CHAPTER 1: Literature Review

The study of personality examines individual differences relating to dispositional factors and associated behaviors that differentiate people from one another (Kluckholm & Murray, 1953). That is, the study of patterns of thoughts, feelings, and behavior unique to an individual. The world personality, derived from the Latin persona, refers to the Greek masks worn by actors to help the audience identify which character, or personality, the actor was portraying. Although personality broadly refers to a person's characteristics, it is suggested that personality may be better understood as patterns of expression for others to see (Hogan & Foster, 2016).

With the increased use of personality assessments in high-stakes contexts such as job selection (Carless, 2007) the extent to which people can “fake” (distort their responses to create the desired impression) and the extent to which faking negatively impacts hiring decisions has been debated (Dilchert et al., 2006; Hough & Oswald, 2008; Zickar & Robie, 1999). Where faking in a low-stakes context may be generally less consequential, faking in a high-stakes context may be the difference between getting or not getting a job (Rosse et al., 1998). The large body of faking research has typically focused on self-report ratings, mainly within the Big Five model of personality assessment (MacCann, 2013). The principal aim of this thesis is to extend this area of research in two ways. First, I will examine the Dark Triad assessment of personality (Paulhus & Williams, 2002). Given the increased use of Dark Triad assessments in selection settings (Spain et al., 2014) and their existing use in forensic settings (Bogaerts et al., 2012; Kavish et al., 2019; Polaschek & Daly, 2013), it is crucial to have a clear understanding of whether people can fake on these measures, and if so, to what extent. Second, I will consider both self-rating and informant-ratings when considering response distortion.
In the paragraphs below, I summarize four key areas of research. First, I will review self-report personality measurement. Second, I will review response distortion and instructed faking, including: a) whether people can fake, b) whether people do fake, c) the impact of faking on the validity of personality measures, and d) the use of informant-reports as an additional measurement method. Third, I review the impact of the evaluative content of items on personality questionnaires and whether self-interested motivations impact the extent to which people may fake for themselves or others. Fourth, I summarize research on the Dark Triad of personality domains and facets.

1.1 Self-Report Measures of Personality

Personality is most commonly measured using self-report rating scales. After all, who better to tell you about their personality than the person themselves? Personality scores are derived by collecting responses to items endorsed by the test-takers based on their self-perceptions (e.g., "I am a hard worker" Strongly Agree, Agree, Disagree, Strongly Disagree). Self-report measures can be completed online or by pen and paper, thereby representing a quick and affordable way to collect information about a person ranging from their beliefs and attitudes to behavioral idiosyncrasies (Olino & Klein, 2015).

Self-reported personality assessment is associated with a variety of social and behavioural outcomes (Muris et al., 2017; Ozer & Reise, 1994). Unsurprisingly, self-report personality tests are used in various contexts such as personnel selection, online dating, and selection for highly competitive educational programs (Ellison et al., 2006; Hojat et al., 2013; Oh et al., 2011). In fact, meta-analytic findings generally provide support for self-reported personality as valid predictors of a variety of life outcomes, including workplace performance, academic performance, and wellbeing (Anglim et al., 2020; Barrick & Mount, 1991; Poropat, 2009).
1.2 Response Distortion

There has been a great deal of debate surrounding the impact of response distortion on the interpretation of personality scores (Birkeland et al., 2006; Wetzel et al., 2016; Zickar & Robie, 1999). The accuracy of self-report rating scales depends on honest responses combined with accurate self-perception (Paulhus, & Vazire, 2007). Not all test-takers respond accurately, though inaccuracy is not necessarily intentional. It may be the case that some individuals prefer to answer accurately to avoid unnecessary expectations (i.e., low conscientious people may be disinclined to feign conscientiousness if it might mean additional workload or responsibility being added to their job). Though it may also be the case that people may intentionally distort their responses when presented with sufficient motivation such as a high-stakes employment context (impression management; Paulhus, 1984, 2002), while others may unconsciously distort their responses (self-deception; Edwards, 1957; Paulhus, 1984, 2002; Paulhus & Vazire, 2007). That is, not all socially desirable responding is deliberate, and for some individuals it may simply occur as a function of self-deceptive enhancement (based on egoistic biases) or self-deceptive denial (based on moralistic biases; Paulhus, 2002).

Whether unconscious or intentional, socially desirable, or undesirable responding can be motivated by psychological factors (maintaining or enhancing an individual’s ego and self-esteem; Paulhus & John, 1998), or instrumental factors (e.g., faking high on a personality test in order to get a well-paying job). The term ‘faking’ is used to indicate intentional response distortion. Faking can involve both: a) socially desirable responding, where test-takers try to make themselves look better than they really are (faking good); and/or or b) socially undesirable responding, where test-takers try to make themselves look worse than they really are (faking bad, referred to as malingering, Rogers et al., 2003). Individuals may fake when the outcome of an assessment is a consequential decision about them (e.g.,
whether they get a job, whether they get into medical school; Anglim et al., 2018; Morgeson et al., 2007). When the interpretation of a test score gives rise to a valued outcome for the individual, the test is said to be “high stakes” (e.g., whether a person gets a job or is admitted into medical school is at stake). When the outcome of a test score interpretation is not critically important (e.g., an individual completing documentation as a job incumbent), the test is said to be low-stakes (Anglim et al., 2018). People are much more motivated to intentionally distort their responses in “high-stakes” as compared to “low-stakes” scenarios (Anglim et al., 2018; Birkeland et al., 2006). Accordingly, the feasibility of self-report scales to present an accurate personality profile in high-stakes settings, has been questioned (Hartman & Grubb, 2011; Zickar & Robie, 1999). Given these concerns, researchers continue to evaluate the psychometric properties of self-report measures and the impact of response distortion on how personality scales are interpreted.

This thesis will specifically examine intentional response distortion (i.e., faking). The potential for response distortion has raised several questions, many of which have been addressed in the literature. First, can test-takers distort their responses on self-report questionnaires (Viswesvaran & Ones, 1999). Second, do test-takers distort their responses on self-report questionnaires (Birkeland et al., 2006; Griffith & Converse, 2012). Third, what is the impact on the interpretation of self-report measures when test-takers distort their responses (Komar et al., 2008; Meston et al., 1998). Finally, what other methods can be employed to get information about a person's personality (Holden & Marjanovic, 2019; Vazire, 2006).

**Instructed Faking: can people fake?**

Two early questions arising in this research area are whether people can fake on personality measures and how much they can fake on these measures. Instructed faking paradigms are a standard method used by researchers to simulate high-stakes settings in order
to determine how much test-takers can distort their responses. Instructed faking paradigm instructions are typically referred to as *faking good* and *faking bad* depending on whether situation would simulate a scenario encouraging the test taker to make themselves look better (fake good) or worse (fake bad) than they really are (Arthur et al., 2010; Donaldson & Grant-Vallone, 2002; Furnham, 1990). For example, a job candidate may try to make themselves look like a better fit for the job than they actually are (Birkeland et al., 2006). Equally, test-takers may *fake-bad* presenting themselves unfavorably. For example, a test-taker may inflate mental illness symptoms in a forensic setting when seeking a lesser sentence or when being assessed for competency to stand trial (Mills & Kroner, 2005). It is important to note that faking on personality tests is due to both *person factors* (stable traits that underlie the propensity to fake) and *situation factors* (the stakes of the test, and the characteristics of the test items).

Instructions rarely instruct participants to lie outright. Instead, instructions usually involve a “real-world” cover story (such as job selection) where an individual simulates a high-stakes response (Mueller-Hanson et al., 2003). It has been recommended that instructed faking studies should specify the job type as different types of jobs may elicit different response profiles (Furnham, 1990; Tett et al., 2012). For research interested in how people fake in a selection context, instructions such as “Imagine you are taking this test as part of the job selection process for a job as a sales director” may be used. In a high-stakes forensic context, instructions such as “Imagine you are taking this test as part of a prison sentencing procedure that will determine whether you can be admitted to a low security prison” may be used (Edens et al., 2001).

Numerous instructed faking studies have investigated the extent to which test-takers fake on personality measures. Typically, instructed faking personality research has focused on the Big-Five personality traits (MacCann, 2013) with a large literature demonstrating that
motivated individuals can fake their personality scores (Viswesvaran & Ones 1999). Deliberate attempts to manipulate one's test result to produce different scores than if they were answering honestly should not be confused with unintentional score inflation due to a lack of self-insight (Paulhus & John, 1998). A meta-analysis of 51 studies found people could successfully inflate their Big-Five personality scores when instructed to fake good compared to answering honestly with differences for both between- and within-subject designs. Faking good effect sizes (Cohen’s $d$) were moderate to large, ranging from 0.48 (agreeableness) to 0.65 (openness) for between-subjects designs, and 0.47 (agreeableness) to 0.93 (emotional stability) for within-subjects designs. (Viswesvaran & Ones, 1999). Faking bad effect sizes were considerably larger (large to extremely large), ranging from -1.0 (emotional stability) to -1.95 (extraversion) for between-subjects designs and -0.91 (extraversion) to -3.34 (emotional stability) for within-subjects designs. Responses obtained during instructed faking studies indicate the upper limits of faking, indicating the extent to which people can fake on these measures.

Instructed faking paradigms are often used in organizational psychology research to investigate faking during the selection process. As such, much of the research about instructed faking is focused on job applicants. Furnham (1990) found that participants produced different personality profiles depending on which 'job' they were applying for. For example, participants' personality profiles for a job as a banker were significantly different from those instructed to fake-good for a librarian or advertising executive position. In similar studies, these personality profiles closely resembled the descriptions of a perfect job candidate (Martin et al., 2002) or archetypal employees (Levashina & Campion, 2006) provided by the participants describing the target job. Prior research gives ample support to the conclusion that test-takers can fake on self-report measures when instructed to do so (Viswesvaran & Ones 1999).
Faking: do people fake?

Instructed faking paradigms are used to examine the extent to which test-takers deliberately attempt to manipulate their test results in a variety of settings including job selection, admission to education programs, and forensic settings. Typically, however, instructed faking is conducted with reference to a job selection context where a person’s Big Five score impacts the likelihood of their selection. Still, instructed faking paradigms do not tell us whether people do fake when given the opportunity to do so. A criticism of instructed faking paradigms is whether they are useful outside the context of a lab environment. Some researchers suggest very little faking occurs in real-life job applicant settings (Dunnette et al., 1962; Hough, et al., 1990). Others have found personality measures are particularly susceptible to faking with higher mean personality scores reported among job applicants than job incumbents (Birkeland et al., 2006; Goffin & Boyd, 2009). Even though there is an ongoing debate regarding criterion-related validity, it is likely that faking negatively impacts hiring decisions. Comparing job incumbents to job applicants presents an opportunity to examine how much people do fake in practice. However, there are limitations to this approach given that such opportunities to compare job applicants and job incumbents are rare in high-stakes situations (Griffith et al., 2007). Nevertheless, meta-analytic results show small (but still substantial) score inflation for job applicants compared to job incumbents on Big-Five measures. These mean standardized difference scores range from 0.11 (extraversion) to 0.45 (conscientiousness; Birkeland et al., 2006). These prior findings provide a basis to support the notion that at least some people fake during the job application process.

There have been several questions raised as to the usefulness of applicant-incumbent comparisons (Hough et al., 1990). These questions generally relate to the motivation or carelessness of job incumbent-reporting. Even in extremely competitive job markets, applicant-incumbent studies have not produced large effects, therefore questioning the utility
of exploring applicant-incumbent differences (Ones & Viswesvaran, 2001). Researchers
could address these concerns by re-administering the test to job applicants following the
application process (Griffith et al., 2007). Using this design methodology, Griffith et al.,
found between 30-50% of job applicants do fake personality assessments during the selection
process. Furthermore, faking impacted the rank order of test-takers such that it significantly
impacted hiring decisions. Research which clearly answers whether test-takers do distort their
responses in real-world, high-stakes settings is ongoing.

The effect of faking on self-report personality measures

There is a large body of research investigating the extent to which the criterion-related
and construct validity of self-reports are diminished due to faking (Ellingson et al., 1999;
Ziegler, 2011). Understanding the impact of faking is especially important when considering
personality scales are used to predict job performance and organizational fit in job applicants
(Barrick & Mount, 1991). To date, there are mixed findings in this area. Drasgow and Kang
(1984) found rank order did not change as a result of response distortion.

In contrast, Griffith et al., (2007) found that rank order was substantially impacted by
response distortion following the study of job applicants' personality scores during and after
the selection process. Several other studies have also found faking on self-report measures
may negatively impact the interpretation of personality scores. Numerous studies have
demonstrated rank order changes that directly affect hiring decisions (Douglas, 1996; Rosse
et al., 1998; Zickar & Robie, 1996). Despite the need for further substantial research in this
area, there is compelling evidence to encourage the continued identification of alternate
methods to accompany self-report scales to improve the reliability of self-reported
personality scores.

1.3 Informant-Report Measures of Personality
While self-report measures remain the most direct route to accessing information about an individual, the use of informant-reports are an alternative way to assess personality. It has been proposed that informant-report ratings can help mitigate the social desirability biases commonly found in self-reported personality scores (Funder & West, 1993). In fact, some researchers suggest that informant-report ratings are preferable to self-report scales due to their ‘objectivity’ (e.g., Kenny et al., 1994). A high degree of similarity between self- and informant-reported scores are said to indicate lower levels of systematic error such as error as a result of socially desirable responding on self-report. The degree to which self-reported personality scores converge with informant-reported personality is known as self-other agreement (Connelly & Ones, 2010; Kim et al., 2014; Vazire, 2010). Meta-analytic findings show considerable self-other agreement on personality measures (Connelly, & Ones, 2010; Kim et al., 2014). However, self-other agreement differs based on the informant's relationship with the target, with stronger agreement when informants and targets know each other well (Connelly, & Ones, 2010; Funder et al., 1995; Harkness et al., 1995).

Several factors may impact the accuracy of informant reporting including characteristics of the informant, the trait being observed, and how observable the trait is to the informant (Funder, 1995). Trait visibility is understood as the extent to which traits are observable or visible (e.g., extraversion is more observable than openness to experience). In such instances, informants not well acquainted with their target may not notice characteristics or behaviors that are less observable (Funder & Dobroth, 1987). The opposite may also be true; informants may observe some traits that otherwise remain unknown to the self (McAbee & Connelly, 2016).

Both Funder and Dobroth (1987), and Nederstrom and Selmela-Aro (2014) demonstrated that self-other agreement is greatest for more visible personality traits (e.g., extraversion) compared to less visible (e.g., neuroticism). Vazire (2010) proposed both
observability and the evaluative nature of item content can impact the accuracy of self- and informant-report ratings. That is, an informant should more accurately report observable personality traits (such as extraversion) compared to less visible personality traits (such as neuroticism; (Vazire, 2010). It is also important to consider that informants are generally selected by the target, in both high- and low-stakes contexts, and are generally people both who know and like the target relatively well (Kolar et al., 1996; Vazire, 2010; Vazire & Carlson, 2010).

**Can Informants Fake for their Target?**

There is a sizeable extant literature dedicated to faking on self-reported measures of personality. However, as yet, faking by informant-reports has not been examined with the same scrutiny as self-reports. How well an informant knows the target and how much the informant likes the target impact the informant's reporting (Hollander, 1956; Leising et al., 2010). Early users of informant-reports raised concerns about potential "friendship effects," suggesting informants with a close relationship to the target may respond favorably about them (Hollander, 1956). These initial concerns were often dismissed with the explanation that informants lack motivation for response distortion as the consequences of response distortion do not directly impact them. More succinctly, informants are not susceptible to the same ego-protective biases as self-report (Vazire, 2010). However, Leising et al., (2010) and Leising et al., (2013) demonstrated that the degree to which an informant likes the target inevitably affects their rating of the target. This is especially important from a selection perspective where the target often chooses who a prospective employer should contact for information about them. That is, it is the target who chooses their informant. If they choose someone who knows, and likes them rather well, then as per Leising et al’s findings, this is likely to impact how the informant rates their target. While these findings suggest informants may be susceptible to inflating their target's positive characteristics, as yet, no study has directly
examined whether informants fake for their target. The implication of faking on informant-reports is far-reaching, especially considering they are often used to assess the accuracy of self-reported personality scores and may be an integral factor in hiring decisions. If, as this thesis aims to investigate, informants can fake on behalf of their target, it is essential to understand what may underlie their motivation.

1.4 Evaluative and Descriptive Item Content

Personality questionnaires contain both descriptive (i.e., content relating to the construct being measured) and evaluative (i.e., the extent to which content sounds good versus bad) content. Anglim et al., (2017) and Bäckström et al., (2009) suggest faking involves individuals making item level evaluations of social desirability. That is, individuals may decide how much to endorse an item based on how good or bad the item sounds to them.

It is not a new observation that the inclusion of evaluation items content in personality inventories is problematic. As early as the 1950s, Edwards (1957) demonstrated that expert ratings of an item’s social desirability were almost perfectly correlated with the personality inventories themselves. The endorsement of positively-valanced items and denial of negatively-valanced items suggests the shared evaluative content (i.e., positive or negative valance) may be driving covariance between items and measures rather than the descriptive content of the items. After controlling for the evaluative item content, Bäckström (2007) found significantly smaller correlations among the Big-Five domains. Further, when the Big-Five items were phrased neutrally, the covariance between items was substantially reduced. Essentially, Bäckström et al. (2009) concluded neutral items elicited less socially desirable responding than positively or negatively valanced evaluative items. Research examining the impact of evaluative content on item endorsement has been conducted almost exclusively within the Big Five and HEXACO personality frameworks with very little examination of the evaluative nature of items in other personality domains such as the Dark Triad of personality.
1.5 Self-interest and other-interest

Endorsing items on self-report questionnaires to make oneself appear better than one is to attain a desired outcome is congruent with the definition of self-interested behaviour. Self-interested behaviour is generally understood as attitudes and behaviours that produce positive personal benefits (Gerbasi & Prentice, 2013). It is generally accepted that individuals are likely to act with regard for their own self-interest (Holmes et al., 2002), that they expect others to act in a self-interested way (Miller & Ratner, 1998), and that these motives can explain a person's behavior (Miller, 1999). Gerbasi and Prentice have identified three situations in which self-interested behavior may emerge: (1) When the cost/benefit outcomes are clear; (2) when it is clear how the costs/benefits will personally impact them; and (3) when they deem self-interest more important than thinking of others (Gerbasi & Prentice, 2013). This implies that self-interested behavior may be deliberate, requiring the individual to explicitly analyze cause and effect within situations to act in their own self-interest. Despite this, Gerbasi and Prentice suggest that despite the necessary deliberate aspects of self-interested behavior, the extent to which individuals embrace their own self-interest is the critical feature.

Individuals taking part in economic game research tend to make decisions inconsistent with self-interested behavior, instead showing concern for fairness and reciprocity (Carmerer & Thaler, 1995; Thaler, 1988). Such concerns for fairness and reciprocity are both self- and other-interested though it seems individuals want others to do well, but not necessarily better than themselves. Where the cost is low, but the reward high, engaging in other-interested behavior is beneficial. This mix of self- and other-interested behaviour may be a factor in whether informants decide to intentionally distort their responses on behalf of their target. Where an informant may rate faking for someone else as low-risk to themselves, they may see it as beneficial to building, or maintaining social
standing with their target. That is, even small social gains may be incentive enough to fake for another person. Social norms and social responsibility, along with high agreeableness, tend to be factors in whether individuals will behave in ways that will benefit others (Graziano et al., 2007).

1.6 The Dark Triad

The 'dark triad' is an umbrella term used to identify a set of three socially aversive personality traits. The dark triad comprises narcissism, Machiavellianism, and psychopathy (Paulhus & Williams, 2002). These three traits are associated with ethical, moral, and socially deviant behavior with a callous and antagonistic core, although they are considered distinct constructs. In fact, even though the dark triad promotes simultaneous measurement of narcissism, Machiavellianism, and psychopathy, each construct has long been empirically studied in its own right.

Narcissism

Narcissism originates from Greek mythologies tale of Narcissus. Narcissus was a young male hunter so consumed by the beauty of his own reflection that he shunned other people's love and attention. Traditionally defined as a personality disorder, the introduction of sub-clinical narcissism into the personality literature followed the publication of the Narcissistic Personality Inventory (NPI; Raskin & Hall, 1979). Characteristics of sub-clinical narcissism are found in the wider population ranging from mild to extreme (Raskin & Hall, 1979; Samuel & Widiger, 2008) and narcissism is commonly examined as a personality trait (Miller & Campbell, 2008). Narcissism is broadly understood as behavioural patterns and attitudes associated with grandiosity, entitlement, and desire for admiration (Krizan & Herlache, 2018). More succinctly, Krizan, and Herlache defined narcissism as possessing 'entitled self-importance.' Additionally, narcissism has been characterized by a desire to
present a socially desirable impression with a tendency toward interpersonal manipulation (Dickinson, & Pincus, 2003).

Narcissism is comprised of two separate but related factors: grandiose and vulnerable narcissism (Besser & Priel, 2010; Krizan & Herlache, 2018; Miller et al., 2011; Rohmann et al., 2012; Wink, 1991). Despite sharing a number of central characteristics such as an antagonistic interpersonal style, entitlement, hypersensitivity to criticism, and self-absorption (Dickinson, & Pincus, 2003; Krizan & Herlache, 2018; Rauthmann, & Kolar, 2012; Weiss et al., 2019), there is clear evidence grandiose and vulnerable narcissism are distinct (Krizan & Herlache, 2018; Miller et al., 2019). Furthermore, even though grandiose and vulnerable narcissism may share several characteristics, the manifestation of these characteristics may differ between the facets.

**Grandiose narcissism.** Individuals high in grandiose narcissism are characterized by grandiosity, self-confidence, exploitation of others, and rely on self/internal validation to maintain self-esteem. Individuals may possess an inflated self-image while simultaneously repressing information to the contrary (Campbell & Foster, 2007). Generally, individuals with high grandiose narcissism scores tend to make an impression on those around them with their extroverted, self-assuredness, and need for admiration (Wink, 1991). When threatened, individuals high in grandiose narcissism tend to blame and devalue others while denying their own weaknesses (Dickinson & Pincus, 2003; Zhang et al., 2017).

**Vulnerable narcissism.** Grandiose fantasies are also characteristic of vulnerable narcissism although individuals with high levels of vulnerable narcissism tend to oscillate between self-love and self-loathing, typically exhibiting a fragile sense of self, along with defensiveness, and resentment (Weiss et al., 2019). Feeling undervalued may lead to withdrawing from interpersonal relationships. Individuals high in vulnerable narcissism can be hypersensitive to negative feedback and act aggressively when their sense of self is
threatened (Wink, 1991). Such individuals also tend to have a fatalistic worldview (Zajenkowski et al., 2016). Vulnerable narcissism tends to rely on external validation to maintain self-esteem, in contrast to grandiose narcissism.

**Machiavellianism**

Machiavellianism is characterized by goal-focused manipulative, duplicitous interpersonal relations and strategic long-term planning (Christie & Geis, 1970). Individuals high in Machiavellianism tend to be viewed as strategic, capable of delaying gratification for bigger and better rewards in the future, possessing low moral commitment, and engaging in long-term strategic planning with a cynical worldview (Christie & Geis, 1970; Furnham et al., 2013). Christie and Geis (1970) proposed Machiavellianism as a personality trait following from the writing of Niccolo Machiavelli, a political advisor to the Medici family in 16th century Firenze. Machiavelli’s seminal work, *Il Principe* ("The Prince"; 1513) contained advice to rulers about the successful governance of society, particularly through the use of others' fortune and power. Specifically, Machiavelli emphasises that using manipulation and cunning leads to success more so than truthfulness and empathy.

**Psychopathy**

Psychopathy has been characterised by interpersonal, affective, and behavioral characteristics, including superficial charm, pathological lying, and lack of empathy, conscience, and remorse (Cleckley, 1951; Hare, 2003). Often confused with antisocial personality disorder, the sub-clinical characteristics of psychopathy exist on a spectrum in the wider population (Berg et al., 2013). Based on Karpman's (1941) work, psychopathy distinguishes between two facets differing in their etiology, influences and symptomology. Expanding on Karpman’s earlier work, Hare (2003) developed a model of psychopathy differentiating *primary* and *secondary* psychopathy. Both psychopathy facets involve affective elements suggesting some indifference to their own and others' emotions, each
underpinned by an antagonistic interpersonal style (Miller & Lynam, 2012). Although people may have both primary and secondary psychopathy manifestations, one facet tends to dominate, providing a more accurate description of a person’s pattern of psychopathic personality (Karpman, 1941).

**Primary psychopathy.** Primary psychopathy is characterized by (a) lacking guilt and remorse, with higher levels of callousness, manipulation, and a desire to maintain social desirability (Hare, 2003), (b) having superficial affect (Casey et al., 2013), (c) deficits in affective empathy (Wai & Tiliopoulos, 2012), and (d) lower levels of fear (Patrick et al., 1993). According to Cleckely (1976), primary psychopathy stems from affective deficits making it difficult to build and maintain interpersonal relationships. In contrast, Lykken (1995) suggested that the underlying core mechanism of primary psychopathy was fearlessness.

**Secondary psychopathy.** Secondary psychopathy is typified by elevated levels of antisocial behaviors such as aggressiveness, impulsivity, and neuroticism (Levenson et al., 1995). These characteristics result from difficulty to effectively regulate the experience of strong emotions (Hare, 2003). Additionally, individuals scoring high on secondary psychopathy possess guilt and fear responses not usually observed in individuals with high primary psychopathy levels (Lykken, 1995; Wallace et al., 2009). Consequently, high secondary psychopathy levels are considered to have a reactive interpersonal style, leading to strong negative affect and aggression (Karpman, 1955).

There are other variations between primary and secondary psychopathy. While primary psychopathy is considered affective and interpersonal (Cleckley, 1951), secondary psychopathy is likely to engage in aggressive outbursts influenced by antisocial and impulsive behaviour (Hare, 2003). Where individuals with higher levels of primary psychopathy tend to have a goal-directed interest in maintaining social desirability,
individuals higher in secondary psychopathy tend not to be interested in maintaining social desirability (Hare 2003). Additionally, prior research has distinguished between primary and secondary psychopathy by demonstrating differential performance deficits relating to high and low anxiety in individuals exhibiting higher psychopathy levels. Specifically, this research suggests that individuals with high levels of primary psychopathy tend to exhibit lower anxiety levels than those with higher levels of secondary psychopathy (Cleckely, 1976; Lykken 1995).

Measuring the dark triad

Interest in the dark triad continues to grow as researchers begin to examine the darker aspects of personality beyond the Big-Five and HEXACO models (Muris et al., 2017; Vize et al., 2018). High scores on the dark triad are said to predict a variety of outcomes related to workplace behavior, leadership qualities, romantic relationship success, and mental health outcomes (for a review, see Muris et al., 2017). Consequently, interest in the dark triad has gone beyond personality research, expanding into criminology (Wright et al., 2017), personnel selection (Spain et al., 2014), and workplace behavior (Forsyth et al., 2012).

As a consequence of growing interest in dark triad assessment, there is an ongoing debate about the most appropriate way to measure these traits at the subclinical level (Miller et al., 2019; Muris et al., 2017). As yet, only two scales have been developed which purport to measure the dark triad as a whole: (1) The Short Dark Triad (SD3: Jones & Paulhus, 2014), and (2) the Dirty Dozen (DD: Jonason & Webster, 2010). The majority of research in this area continues to use standard self-report measures developed specifically for each of the traits such as the Narcissistic Personality Inventory (Raskin & Hall, 1979), the Mach IV (Christie & Geis, 1970), and the Levenson's Self-Report Scale (Levenson et al., 1995) among others (Miller et al., 2019). As self-report measures rely on the assumption of truthful responding and the individual's introspective ability to answer truthfully (Paulhus & Vazire,
2007), it is essential to understand the extent to which these assumptions are accurate for individuals when given an opportunity to fake. Furthermore, examining the degree to which individuals are influenced by the evaluative nature of the items in both self- and informant-reports may provide an indication of whether self-other agreement demonstrates a lack of self-reported socially desirable responding as believed, or instead demonstrates that informants are also influenced by evaluative content.

1.7 Proposed research on measuring faking on self- and informant-reports of the dark triad

Given the complexity involved in measuring the fakeability of dark triad scales, combined with a lack of scrutiny relating to faking on informant-report measures of personality, the current thesis examines faking on self-report measures of the dark triad (Chapter 2) followed by an investigation of faking on both self- and informant-report measures of the dark triad (Chapter 3).

1.8 Overview and Aims of the Present Thesis

This thesis aims to address two critical issues identified in the theoretical and methodological considerations raised in this introductory chapter:

1) The extent to which self- and informant-report measures of the dark triad can be faked.

2) The extent to which people can fake-good and fake-bad on measures of the dark triad.

Investigating these issues is critical to clarify the extent to which self- and informant-report measures of the dark triad can be relied upon to provide an accurate assessment of these traits. This thesis will focus on the dark triad rather than the Big-Five or HEXACO models of personality. Faking on the dark triad is arguably more consequential than faking on other measures of personality. Arguably, errors related to releasing offenders into the
community, or training law enforcement officers, or medical practitioners with high levels of psychopathic traits may be more consequential than failing to select the best candidate for a managerial position. Given the increased use of Dark Triad assessments in selection settings (Spain et al., 2014) and their existing use in forensic settings (Bogaerts et al., 2012; Kavish et al., 2019; Polaschek & Daly, 2013), it is crucial to have a clear understanding of whether people can fake on these measures, and if so, to what extent.

The meta-analysis and systematic review presented in Chapter 2 addresses the thesis's first aim through the systematic review of prior instructed faking studies using the dark triad measures. Although an extensive literature has examined social desirability alongside dark triad assessment, very little research has addressed whether people can fake on dark triad measures. Chapter 2 illustrates that similar to related Big-Five meta-analyses (e.g., Viswesvaran & Ones, 1999), faking on dark triad measures is substantial. People fake-bad to a greater extent than they fake-good. This meta-analysis highlights a concerning dearth of research in this area, particularly given the implications of response distortion on dark triad measures across personality, forensic, organizational, and social psychological research.

Study 1 (Chapter 3) investigates the extent to which people can fake self- and informant-report measures of the dark triad when instructed to do so. This study uses a 2x3 between-subjects design, where participants are randomly assigned to one of six potential conditions where instructions (answer honestly, fake-good, or fake-bad) were crossed with target (self-report, or informant-report). Participants will also complete a self- and other-interest questionnaire to explore whether these attitudes influence the extent to which people fake on self-reports compared to informant-reports. Study 2 (Chapter 3) examines the extent to which faking on Dark Triad measures is influenced by the evaluative content of the scales.

The contents of this chapter are a minor revision of the following submitted manuscript


2.1 Abstract

Prior meta-analyses demonstrate that people can intentionally distort Big Five personality scores if instructed to do so. As yet, there is no equivalent meta-analysis addressing instructed faking on the dark triad (narcissism, Machiavellianism, and psychopathy). Therefore, we review mean score changes to the dark triad domains and facets under instructed faking. Due to insufficient $k$ for meta-analysis, narcissism and Machiavellianism were systematically reviewed alongside psychopathy. Surprisingly, the single study examining ‘faking good’ on narcissism showed a large increase. In contrast, Machiavellianism scores decreased under ‘fake good’ instructions but increased under ‘fake bad’ instructions. The psychopathy meta-analyses showed that: a) scores significantly decreased under ‘fake good’ instructions ($k = 10$, $N = 1,638$) for total, primary and secondary psychopathy ($d = -0.62$, -0.84 and -0.58); and b) scores significantly increased under ‘fake bad’ instructions ($k = 5$, $N = 629$) for total, primary and secondary psychopathy ($d = 1.81$, 0.90 and 1.08). For total psychopathy, the magnitude of faking was significantly larger for ‘fake bad’ versus ‘fake good.’ We conclude that dark triad measures are fakeable to a similar extent as the Big Five. We discuss the relevance of our findings for psychopathy and dark triad assessment in several applied contexts.

Keywords: psychopathy, meta-analysis, faking, response distortion, personality
2.2 Introduction

People are highly motivated to be viewed in a favorable light by others. There is evidence that people distort their responses on personality scales under high-stakes conditions (Sjoberg, 2015). In fact, meta-analytic findings show people can and do substantially change their personality scores when motivated to do so (Birkeland et al., 2006; Viswesvaran & Ones, 1999). Much of this research has focused on the broad personality domains of the Big Five/Five-factor model (extraversion, agreeableness, conscientiousness, neuroticism, and openness; MacCann, 2013). As yet, no meta-analysis has addressed whether people also fake on the dark triad of personality (narcissism, Machiavellianism, and psychopathy; Paulhus & Williams, 2002). The current study addresses this gap by conducting a systematic review and meta-analysis examining the extent to which people can distort their responses on the dark triad domains and facets when instructed to do so. While domains, facets and instruments of the dark triad differ, dark triad items generally represent ‘dark’ content (and therefore sound undesirable). As such, we expect response distortion will be at least as marked for the dark triad as the Big Five.

Response Distortion

Personality traits are commonly measured using self-report rating scales (e.g., “I am a hard worker” Strongly Agree, Agree, Disagree, Strongly Disagree). The accuracy of self-report rating scales relies on truthful responding combined with the individual’s introspective ability (Paulhus, & Vazire, 2007). However, test-takers may intentionally distort their responses when motivated to do so in order to achieve a specific goal (Tett & Simonet, 2011). They may present themselves more favorably than they really are (faking good) or less favorably than they really are (faking bad, also referred to as malingering; Rogers et al., 2003). This intentional response distortion is referred to as ‘faking’, and there is evidence that it occurs when the outcome of the test scores are consequential for the test-taker. For
example, people taking personality tests as part of job selection processes tend to report more positive scores than employees taking personality tests for non-selection purposes (Birkeland et al., 2006). As a conservative estimate, 30-50% of job applicants do not represent themselves accurately on self-report measures (Griffith & Converse, 2012). These prior findings suggest that people do fake, though not all people fake to the maximum possible extent all the time (Birkeland et al., 2006; Griffith & Converse, 2012; Viswesvaran & Ones, 1999).

**Instructed Faking**

Self-report measures have been criticized for their susceptibility to deliberate distortion (Griffith et al., 2007). A standard method used to determine the extent to which self-report measures can be distorted is instructed faking. Instructed faking is a paradigm by which individuals are provided with a set of instructions to answer honestly or distort their responses to manage others’ perceptions of them. To distort their responses, test-takers are instructed to either fake-good or fake-bad. Faking-good occurs when people try to make themselves look better than they are (Arthur et al., 2010; Donaldson & Grant-Vallone, 2002; Furnham, 1990). Conversely, test-takers may fake-bad to make themselves look worse than they are. For example, within a forensic setting, a test-taker may inflate symptoms when seeking reparation for psychological damage or when an offender is being assessed for competency to stand trial (Mills & Kroner, 2005).

Despite being a common paradigm to measure faking there are no established gold-standard guidelines of how to best elicit faking in participants. In earlier instructed faking paradigms, test-takers were instructed to answer in a way that makes them an ideal candidate for a job they want. With each test-taker presumably choosing a wide range of different target jobs, so too will they produce a wide range of personality profiles potentially limiting the ecological validity of the studies (Furnham, 1990). More recently, there is a focus on
providing specific information relating to the context, and clear instructions for test-takers. Test-takers who have been instructed to present themselves as a good candidate for a specific job such as a librarian, banker, or advertising executive typically produced different personality profiles for each of the jobs (Furnham, 1990). Similarly, when asked to describe an ideal employee, people tended to create response profiles closely aligned to the ideal or stereotypical employee (Levashina & Campion, 2006; Martin et al., 2002). While some research suggests response-distortion in real-world recruitment settings is trivial (Hough et al., 1990), other research reports that job applicants tend to rate themselves more positively on Big-Five domains compared to job incumbents (Birkeland et al., 2006; Griffin et al., 2004). Similar response patterns emerge when using the HEXACO. Individuals in a high-stakes job applicant setting represent themselves significantly more favorably than individuals in a low-stakes setting (Anglim et al., 2017). Collectively, there is evidence suggesting applicants capitalize on the susceptibility of personality rating scales to faking in order to present themselves in a more desirable light. Though it should also be noted that if an individual’s baseline responses are already positively biased, then it may be that the individual has less score variability between ‘honest’ and ‘fake good’ but more variability between ‘honest’ and ‘fake bad’.

Instructed faking paradigms provide the opportunity to examine the extent to which deliberate attempts to manipulate one’s test result produce a different score as compared to answering honestly (without intentional manipulation of one’s score). This is different to the unintentional score inflation that occurs due to lack of self-insight (when people have an unrealistically positive or negative view of themselves; Paulhus & John, 1998). Typically, instructed faking personality research has focused on the Big Five domains (MacCann, 2013) with a large literature highlighting the ability of motivated individuals to fake their scores on personality measures (Viswesvaran & Ones 1999). A meta-analysis of 51 studies found that
people could successfully increase their big five personality scores when instructed to fake good compared to answer honestly with both between-subjects and within-subjects differences. Faking good effect sizes (Cohen’s $d$) were moderate to large, ranging from 0.48 (agreeableness) to 0.65 (openness) for between-subjects designs, and 0.47 (agreeableness) to 0.93 (emotional stability) for within-subjects designs (Viswesvaran & Ones, 1999). Faking bad effect sizes were considerably larger (large to extremely large), ranging from -1.00 (emotional stability) to -1.95 (extraversion) for between-subjects designs and -0.91 (extraversion) to -3.34 (emotional stability) for within-subjects designs. These instructed-faking paradigms are generally thought to capture the extent to which people can fake on self-reported personality scales. In contrast, comparing job incumbents to job applicants indicates how much people actually do fake, in practice. Meta-analyses of this latter paradigm show smaller (but still substantial) big five score changes due to faking, with small to moderate effect sizes, ranging from 0.11 (extraversion) to 0.45 (conscientiousness; Birkeland et al., 2006). Response distortion represents an ongoing challenge for research and practice that relies on self-ratings to gain information about an individual (Komar et al., 2008; Meston et al., 1998).

**The Dark Triad**

The ‘dark triad’ is an umbrella term for a set of three socially aversive personality traits comprised of narcissism, Machiavellianism, and psychopathy (Paulhus & Williams, 2002). While all three traits are associated with ethical, moral, and socially deviant behavior, among other shared characteristics, they are considered independent of each other. Recent debate relating to a shared common core among the dark triad traits continues, but there is some consensus on the role of antagonism connecting narcissism, Machiavellianism, and psychopathy (Jones & Figueredo, 2013; Jones & Neria, 2015; Paulhus & Williams, 2002).
Furthermore, each of the three traits shares exploitative characteristics with goal-focused manipulation of others’ emotions to get what they want.

**Narcissism**

While narcissism has traditionally been conceptualized as a personality disorder, mild to extreme sub-clinical characteristics of narcissism are found in non-clinical populations (Raskin & Hall, 1979; Samuel & Widiger, 2008). Accordingly, narcissism is commonly regarded and investigated as a personality trait (Miller & Campbell, 2008). Narcissism is widely understood as patterns of behaviors and attitudes aligned with grandiosity, self-importance, and desire for admiration (Krizan & Herlach, 2018). More succinctly, Krizan and Herlache defined narcissism as possessing ‘entitled self-importance’. Additionally, narcissism has been characterized by a desire to remain socially desirable and interpersonally manipulative (Dickinson & Pincus, 2003).

Narcissism comprises two distinct but related factors: grandiose and vulnerable narcissism (Besser & Priel, 2010; Krizan & Herlache, 2018; Miller et al., 2011; Rohmann et al., 2012; Wink, 1991). Though grandiose and vulnerable narcissism share several core characteristics such as an antagonistic interpersonal style, self-importance, entitlement, and hypersensitivity to criticism (Dickinson, & Pincus, 2003; Krizan & Herlache, 2018; Weiss et al., 2019), there is clear evidence that these two facets are distinct (Krizan & Herlache, 2018; Miller et al., 2019).

**Grandiose narcissism.** As the name suggests, individuals high in grandiose (or “overt”) narcissism are characterized by grandiosity, self-confidence, and exploitation of others and tend to rely on self-validation to maintain self-esteem. When threatened, individuals high in grandiose narcissism tend to blame and devalue others while denying their own weaknesses (Dickinson & Pincus, 2003; Zhang et al., 2017).
**Vulnerable narcissism.** Vulnerable (or “covert”) narcissism tends to rely on external validation. While grandiose fantasies also characterize vulnerable narcissism, people high on vulnerable narcissism tend to oscillate between self-love and self-loathing, thereby exhibiting fragile self-esteem, defensiveness, and resentment (Weiss et al., 2019). Individuals high in vulnerable narcissism are characteristically hypersensitive to negative feedback and tend to act aggressively when their sense of self is threatened (Wink, 1991).

**Machiavellianism**

Unlike narcissism and psychopathy, Machiavellianism is commonly understood to be a unidimensional construct characterized by goal-focused manipulative and callous social interactions (Christie & Geis, 1970). Individuals high in Machiavellianism tend to be viewed as strategic, capable of delaying gratification for bigger and better rewards in the future, possessing low moral commitment, and engaging in long-term strategic planning with a cold and cynical world-view (Christie & Geis, 1970; Furnham et al., 2013). Machiavellianism, as a personality trait, was derived by Christie and Geis (1970) from the writing of Niccolo Machiavelli, a political advisor to the Medici family in 16th century Firenze. Machiavelli’s ideas were captured in his seminal work, *Il Principe* (“The Prince”; 1513). *Il Principe* contains advice to rulers about successful governance, particularly through the fortune and power of others. Specifically, Machiavelli asserts that the use of manipulation and cunning is more successful than truthfulness.

**Psychopathy**

Psychopathy has been typified by interpersonal, affective, and behavioral characteristics, including superficial charm, pathological lying, and lack of empathy, conscience, and remorse (Cleckley, 1951; Hare, 2003). Often confused with antisocial personality disorder, the sub-clinical characteristics of psychopathy, like narcissism, exist on a continuum in the wider population (Berg et al., 2013). Based on Karpman’s (1941) work,
enduring classifications of psychopathy make a distinction between two subtypes differing in their etiology and symptomology. Expanding on earlier models, Hare (2003) made a distinction between primary and secondary psychopathy. Both psychopathy facets involve affective elements suggesting some indifference to their own and others’ emotions, each underpinned by an antagonistic interpersonal style (Miller & Lynam, 2012).

**Primary psychopathy.** Primary psychopathy is characterized by: (a) a lack of guilt and remorse, with elevated levels of callousness, manipulation, and socially desirable behavior (Hare, 2003), (b) having superficial affect (Casey et al., 2013), and (c) deficits in affective empathy (Wai & Tiliopoulos, 2012). Primary psychopathy is also associated with lower levels of fear (Patrick et al., 1993) and lower indications of repentance (Hare, 2003; Lee & Salekin, 2010).

**Secondary psychopathy.** Secondary psychopathy is characterized by higher levels of antisocial behaviors such as aggressiveness, impulsivity, and anxiety (Levenson et al., 1995). These types of characteristics are potentially a result of experiencing strong emotions which are unable to be effectively regulated (Hare, 2003). Additionally, individuals high on secondary psychopathy have been shown to possess guilt and fear responses not typically observed in individuals high in primary psychopathy (Lykken, 1995; Wallace et al., 2009). It is generally understood psychopathy lies on a continuum with manifestations of both primary and secondary psychopathy. However, one facet tends to dominate providing a better description of an individual’s pattern of psychopathic personality (Karpman, 1941).

There are additional differences between primary and secondary psychopathy. While primary psychopathy is considered affective and interpersonal (Cleckley, 1951) secondary psychopathy is comprised of antisocial and impulsive behaviors with a tendency toward aggressive outbursts (Hare, 2003). Individuals with higher levels of primary psychopathy tend to be goal-directed and interested in maintaining social desirability, whereas individuals
higher in secondary psychopathy are not characterized by the same interest in social desirability maintenance (Hare 2003). Additionally, prior research has attempted to distinguish between primary and secondary psychopathy by showing differential performance deficits relating to high and low anxiety in individuals exhibiting higher levels of psychopathy. Specifically, this research indicates individuals higher in primary psychopathy tend to exhibit lower levels of anxiety compared to those with higher levels of secondary psychopathy (Cleckely, 1976; Lykken 1995).

**Instructed Faking on the Dark Triad**

This meta-analysis will focus on faking on dark triad assessments. In contrast to the Big Five personality traits (where tests are mainly used for job selection), dark triad assessments (particularly psychopathy) are used for forensic applications, including risk assessment for sex offenders but also for ‘screening out’ those with psychologically deviant tendencies from law enforcement and armed services. Arguably, an error in releasing an offender into the community or admitting a psychopathic individual into law enforcement, or even into high ranking governmental positions has more serious consequences than failing to select the best candidate for a graduate or managerial job.

Our expectations regarding the effects of instructed faking are largely based on similar meta-analyses of other personality traits (i.e., the Big Five). There were four broad trends in Viswesvaran and One’s (1999) meta-analysis of instructed faking on the big five. First, there were large effects of instructed faking. Second, results were of moderate to large effect size for all five domains—it did not seem to be the case that faking was limited to specific traits (though there is evidence that people fake particular traits for particular jobs; Tett et al., 2012). Third, effects were slightly larger for within-person than between-person effects. Fourth, effects were clearly larger for ‘fake bad’ than ‘fake good’ studies. This may be due to Big Five items typically being rated well above the rating-scale mid-point (e.g.,
Lang et al., 2011), leaving more room to decrease scores (“fake bad”) than increase scores (“fake good”). For example, mean ratings on conscientiousness items ranged from 5.2 to 6.4 out of 7 in Lang et al., such that scores could maximally increase by 1 to 2 scale-points but maximally decrease by 4 to 5 scale-points.

Because findings were relatively consistent across all five domains, we might assume that the dark triad will similarly show large effects of instructed faking that are relatively similar across all domains. Like the Big Five, ratings on dark triad traits are not centered on the mid-point, but rather towards the desirable end of the scale (lower than the mid-point, for dark triad items). We therefore expect that mean scores on the dark triad would increase under ‘fake bad’ instructions but decrease under ‘fake good’ instructions, with larger effects of ‘faking bad’ than ‘faking good’.

Hypotheses

Hypothesis 1: Instructed faking will result in significant score changes for all three dark triad domains. Compared to ‘answer honestly’ instructions, ‘fake good’ instructions will result in significantly lower mean scores (H1a) whereas ‘fake bad’ instructions will result in significantly higher mean scores (H2a).

Hypothesis 2: Effects for ‘fake bad’ instructions will be of larger magnitude than those for ‘fake good’ instructions.

2.3 Method

Literature Search

A search was conducted in January 2020. The search terms (“dark triad” or “narciss*” or “psychopathy” or “Machiavellian*”) AND (“fak*” or “fake good” or “fake bad” or “instructed faking”) yielded 814 results from the databases PsychInfo, Web of Science, Business Source Ultimate, Medline, and ProQuest Dissertations & Theses, of which titles and abstracts were scanned. The search was limited to English language studies. Additionally,
searches of the reference lists of the identified studies were conducted to identify possible additional studies that may have been missed in the initial search. The database search combined with the reference search yielded a total of 26 studies that progressed to full-text review to check for specific inclusion and exclusion criteria. The overall literature search resulted in 13 studies from 10 articles which fit our inclusion criteria (see Figure 2-1).

**Figure 2.1**

*Quorum Chart of Studies Included in the Systematic Review and Meta-analysis*

![Quorum Chart](image)

**Coding**

The coding procedure was developed based on Cochrane collaboration standards (Higgins & Green, 2011). Study characteristics comprising, author, date, sample size, percentage of female participants, mean age, dark triad instrument, dark triad domains (narcissism, Machiavellianism, psychopathy and related facets), Cronbach’s alpha and Cohen’s $d$ were extracted and coded into a worksheet. Quality control was conducted in order to form a quality index. The first and last author independently double coded all included studies. Any discrepancies were resolved by the first author checking the original manuscript.
In the case of missing data, authors of the applicable study were contacted and invited to send through data for inclusion in the analysis.

**Inclusion and Exclusion Criteria**

Inclusion criteria were: (a) the study (experimental, quasi-experimental, within or between subjects) included a dark triad measure previously published in a publication manual, prior journal article, or academic book chapter; (b) an instructed faking paradigm was used; (c) an honest condition was included as a comparison group; (d) a statistical comparison between honest and faking conditions sufficient to calculate a Cohen’s $d$ estimate; and (e) the paper was written in English.

**Meta-Analytic Approach**

A random effects meta-analysis was carried out to better control for between-study variation (Konstantopoulos, 2006). The primary outcome was standardized mean differences between honest and faking conditions, calculated using Cohen’s $d$. Effect sizes were extracted from studies meeting the eligibility criteria and were used as the measure of effect size in the meta-analysis. Cochran’s $Q$ statistic and $I^2$ were used to evaluate heterogeneity of correlations across included studies (Higgins & Thompson, 2002). Publication bias was evaluated by visually assessing funnel plots, Egger’s test, and Rosenthal’s Fail-safe $N$-test (Egger et al., 1997; Rosenthal, 1991). All analyses were performed using the Metafor package (Viechtbauer, 2010).

Due to the limited number of studies measuring narcissism and Machiavellianism within an instructed faking paradigm a meta-analysis was not appropriate. Therefore, a systematic review was conducted for narcissism and Machiavellianism. Only studies measuring psychopathy were meta-analyzed. There were separate analyses for each of the three domains (and the facets where applicable) for fake-good and fake-bad.

**2.4 Results**
Our search identified 13 studies in total (10 articles). Table 2.1 shows a complete list of all studies included. For narcissism, there was one study, which examined ‘faking good’. For Machiavellianism, there were three studies, one examining ‘faking good’ and two examining ‘faking bad’. For psychopathy, 11 studies examined ‘faking good’ and seven studies examined ‘faking bad’. Of the included studies seven used a student sample, four used a forensic sample (prison inmates), and two used a community sample for a total of 2,169 participants. The mean age range of participants was between 15.21 and 35.98 years. Reviewed studies were grouped into two categories: faking good and faking bad. A summary of the main results obtained with each measure can be found in Table 2.1.

2.4.1 Systematic Review Summary

Narcissism

The sole study measuring the extent to which individuals fake good on narcissism measures compared a multidimensional forced-choice dark triad (Young, 2018). The study was a within-subjects design which instructed participants to fake so as to look like an ideal job applicant in a job selection context. Participants were also informed that that the test-taker with the highest score would receive a $10 bonus. Effect sizes were very large and indicated higher narcissism scores were obtained in ‘fake good’ compared to ‘answer honestly’ instructions. Effects were similar for the two different forced-choice scoring methods ($d = 1.22$ and $d = 1.35$), and for the single-stimulus measure ($d = 1.13$). The mixed community sample consisted of 705 participants with a mean age of 35.98 and was a within-subjects design. The effect was in the opposite direction to our hypothesis, such that Hypothesis 1 was not supported for narcissism.

Machiavellianism

Three studies examined faking on measures of Machiavellianism. The first [Young, 2018] was the study described above for narcissism (i.e., it examined ‘faking good’ in a job
application context on a multidimensional forced choice and single-stimulus administration of a dark triad measure, and promised a $10 bonus for high scores). The study reported significant decreases in mean scores for both the forced-choice version ($d = -0.36$ and $-0.89$ for the two different scoring methods) and the single-stimulus version ($d = -0.28$).

The second study [Skinner et al., 1976] examined the extent to which individuals ‘fake bad’ on Machiavellianism measures using the Mach V (Brodt, 2009) with an effect size of 2.91. This between-person study consisted of 72 participants with a mean age of 19.3. Half of the participants completed the Mach V under standard instructions (“honest” condition). The other half were given a short description of a Machiavellian personality profile followed by instructions to fill in the questionnaire “the way a Machiavellian would fill it in” (Skinner et al., 1976, p. 198; “fake bad” condition). The third study [Skinner, 1982] similarly examined ‘faking bad’ on the Mach V based on instructions to complete the questionnaire like a Machiavellian. Participants were 148 psychology undergraduates who first completed the study under honest conditions, then a week later completed it under ‘fake bad’ conditions. Effects of faking bad were reported only for the 135 participants with Machiavellianism scores in the ‘honest’ condition that were lower than 1 standard deviation above the mean. Mach V scores substantially increased ($d = 1.99$). These results support Hypotheses 1 and 2 for Machiavellianism, in that fake-good instructions resulted in lower Machiavellianism scores (H1a), fake-bad instructions resulted in higher Machiavellianism scores (H1b), and the magnitude of score change was considerably larger for ‘fake bad’ compared to ‘fake good’ instructions.

**Psychopathy**

Thirteen studies examined the extent to which individuals fake on measures of psychopathy. Of these, ten studies assessed faking good [Anderson et al., (2013); Edens (2001; sample 1-3); Edens et al., (2004); Groth et al., (2019); Kelsey et al., (2015); study 1-
2); O’Mahony et al., (1991); Rogers et al., (2002); Young (2019)] with Cohen’s $d$ ranging from -0.05 to -1.93. Six student samples, two mixed community samples, and three forensic samples made up the 1,638 participants with a mean age range of 15.21 to 35.98 years. Of these studies, all were within-subjects design except Anderson et al., which was a between-subjects design. The student and community samples included a mix of job specific and general impression management instructions [Edens (2001); Groth et al., (2019); O’Mahoney et al., (1991); Young (2018)). However, Anderson et al., and Edens asked their student samples to positively manage impressions in an imagined forensic setting. Three forensic studies asked participants to answer in a way to make themselves look better during forensic evaluation. Five studies assessed faking bad [Anderson et al., (2013); Edens et al., (2000); Edens et al., (2004); Groth et al., (2019); Rogers et al., (2002)] with Cohen’s $d$ ranging from 0.10 to 2.65. The three studies using student samples instructed participants to fake a mental illness, the community sample encouraged participants to make themselves look socially undesirable, and the forensic sample instructed participants to appear “tougher” than they really were. A total of 629 participants across the studies comprised three student samples, one mixed community sample, and one forensic sample with a mean age range of 15.21 to 28 years.

2.4.2 Meta-analysis

We examined the standardized mean differences between faking instructions and honest instructions on psychopathy measures using a random-effects meta-analysis. Given that such analyses are unreliable with small samples ($df < 4$), we did not perform an analysis for either narcissism ($n = 1$; Fisher et al., 2017) or Machiavellianism ($n = 3$, Fisher et al., 2017). The demographic and methodological characteristics of the studies included in the meta-analysis are presented in Table 2.1.

Effects of Instructed Faking on Total Psychopathy
**Hypothesis 1: Score changes under instructed faking.** We first examined whether there were significant differences between honest responding compared to instructed faking responding (see Table 2.2). There was an overall significant decrease in psychopathy scores under fake good instructions compared to honest instructions ($d = -0.61$, $SE = 0.20$, $p = .002$). This supports Hypothesis 1a. There was an overall significant increase in psychopathy scores under fake bad instructions compared to honest instructions ($d = 1.81$, $SE = .018$, $p < .001$). This supports Hypothesis 1b.

**Hypothesis 2: Magnitude of ‘fake bad’ score changes is larger than ‘fake good’.** To test whether the magnitude was significantly different for ‘fake good’ versus ‘fake bad’ instructions, we tested whether the 95% confidence intervals for the magnitude overlapped (considering magnitude only, as they were in opposite directions). The confidence intervals for ‘fake good’ (95% C.I: 0.22 to 1.01) and ‘fake bad’ (95% C.I.: 1.46 to 2.17) did not overlap, such that the magnitude of effect is significantly larger for ‘fake bad’ than ‘fake good’, as supporting Hypothesis 2. This difference was also significant when a meta-regression was performed using the absolute value of the effect size as the criterion ($b = 1.191$, $p = .001$).
<table>
<thead>
<tr>
<th>Study</th>
<th>Author(s) &amp; year</th>
<th>Sample Demographics</th>
<th>Measures Dark Triad</th>
<th>Design</th>
<th>N</th>
<th>Mean (SD)</th>
<th>( d )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Participant Profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>% Female</td>
<td>Mean Age</td>
<td>Country</td>
<td></td>
<td>Honest Faking</td>
<td>Honest Faking</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Anderson et al., (2013)</td>
<td>Student</td>
<td>58.0</td>
<td>19.90</td>
<td>USA</td>
<td>PPI</td>
<td>Total Psychopathy</td>
</tr>
<tr>
<td>2.1</td>
<td>Edens (2001)</td>
<td>Student</td>
<td>69.0</td>
<td>21.65</td>
<td>USA</td>
<td>PPI</td>
<td>Total Psychopathy</td>
</tr>
<tr>
<td>2.2</td>
<td>Edens (2001)</td>
<td>Student</td>
<td>69.0</td>
<td>21.65</td>
<td>USA</td>
<td>PPI</td>
<td>Total Psychopathy</td>
</tr>
<tr>
<td>2.3</td>
<td>Edens (2001)</td>
<td>Student</td>
<td>73.0</td>
<td>20.69</td>
<td>USA</td>
<td>PPI</td>
<td>Total Psychopathy</td>
</tr>
<tr>
<td>3</td>
<td>Edens et al., (2004)</td>
<td>Student</td>
<td>73.0</td>
<td>20.69</td>
<td>USA</td>
<td>PPI</td>
<td>Primary</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Groth et al., (2019)</td>
<td>Mixed</td>
<td>56.0</td>
<td>28</td>
<td>Poland</td>
<td>PPI-R</td>
<td>Total Psychopathy</td>
</tr>
<tr>
<td>5</td>
<td>O’Mahony et al., (1991)</td>
<td>Student</td>
<td>0</td>
<td>19.90</td>
<td>USA</td>
<td>SO</td>
<td>Total Psychopathy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Forensic</td>
<td>0</td>
<td>19.40</td>
<td></td>
<td></td>
<td>Total Psychopathy</td>
</tr>
<tr>
<td>6</td>
<td>Young (2019)</td>
<td>Community</td>
<td>66.0</td>
<td>35.98</td>
<td>USA</td>
<td>MFC</td>
<td>Total Psychopathy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Rogers et al. (2002)</td>
<td>Forensic</td>
<td>35.1</td>
<td>15.21</td>
<td>USA</td>
<td>PCL</td>
<td>Primary</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary</td>
<td>39</td>
<td>39</td>
<td>3.42 (1.91)</td>
<td>2.21 (1.44)</td>
<td>0.72 ***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PSD</td>
<td>39</td>
<td>39</td>
<td>8.42 (2.63)</td>
<td>6.44 (2.72)</td>
<td>0.83 ***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRP-II</td>
<td>39</td>
<td>39</td>
<td>105.89 (26.19)</td>
<td>92.26 (19.52)</td>
<td>0.82 ***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Primary</td>
<td>39</td>
<td>39</td>
<td>33.27 (7.35)</td>
<td>34.67 (6.17)</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary</td>
<td>39</td>
<td>39</td>
<td>45.08 (15.40)</td>
<td>36.39 (13.29)</td>
<td>0.86 ***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.1</td>
<td>Kelsey et al., (2015)</td>
<td>Forensic</td>
<td>0</td>
<td>34.06</td>
<td>USA</td>
<td>SRP-4</td>
<td>Total Psychopathy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LSRP</td>
<td>50</td>
<td>50</td>
<td>60.08 (12.38)</td>
<td>44.28 (11.20)</td>
<td>-1.34 ***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Primary</td>
<td>50</td>
<td>50</td>
<td>35.70 (8.46)</td>
<td>25.98 (7.34)</td>
<td>-1.23 ***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary</td>
<td>50</td>
<td>50</td>
<td>24.38 (5.53)</td>
<td>18.3 (5.20)</td>
<td>-1.13 ***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PPI-R</td>
<td>50</td>
<td>50</td>
<td>316.96 (35.39)</td>
<td>264.12 (33.84)</td>
<td>-1.53 ***</td>
</tr>
</tbody>
</table>
### FAKE GOOD

<table>
<thead>
<tr>
<th>Study</th>
<th>Author(s) &amp; year</th>
<th>Sample Demographics</th>
<th>Measures Dark Triad</th>
<th>N</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Participant Profile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>% Female</td>
<td>Mean Age</td>
<td>Country</td>
<td></td>
</tr>
<tr>
<td>8.2</td>
<td>Kelsey et al., (2015)</td>
<td>Forensic</td>
<td>0</td>
<td>34.06</td>
<td>USA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### FAKE BAD

<table>
<thead>
<tr>
<th>Study</th>
<th>Author(s) &amp; year</th>
<th>Sample Demographics</th>
<th>Measures Dark Triad</th>
<th>N</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Anderson et al., (2013)</td>
<td>Student</td>
<td>58.0</td>
<td>19.90</td>
<td>USA</td>
</tr>
<tr>
<td>10</td>
<td>Edens et al. (2000)</td>
<td>Student</td>
<td>68.5</td>
<td>21.65</td>
<td>USA</td>
</tr>
<tr>
<td>3</td>
<td>Edens et al., (2004)</td>
<td>Student</td>
<td>68.0</td>
<td>21.65</td>
<td>USA</td>
</tr>
<tr>
<td>3</td>
<td>Edens et al., (2004)</td>
<td>Student</td>
<td>68.0</td>
<td>21.65</td>
<td>USA</td>
</tr>
<tr>
<td>4</td>
<td>Groth et al., (2019)</td>
<td>Mixed</td>
<td>56.0</td>
<td>28.0</td>
<td>POLAND</td>
</tr>
<tr>
<td>7</td>
<td>Rogers et al. (2002)</td>
<td>Forensic</td>
<td>35.1</td>
<td>15.21</td>
<td>USA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Between = between subjects design; Within = within subjects design; n.p. = not provided; LRSP = Levenson’s Self-Report Psychopathy (Levenson et al., 1995); Mach V = Machiavellianism Scale (Brodt, 2018); MFC = Multidimensional Forced-Choice (Young, 2018); PCL = Psychopathy Checklist (Hare, 1990); PPI = Psychopathic Personality Inventory (Lilienfeld & Andrews, 1996; PPI-R = Psychopathic Personality Inventory Revised (Lilienfeld & Widows, 2005); PSD = Psychopathy Screening Device (Frick & Hare, 2001); SO = Socialization Scale (Gough, 1960); SRP-II = Self-report Psychopathy Scale - II (Hare, 1991); SRP-4 = Self-report Psychopathy Scale 4th Edition (Paulhus et al., 2017).

* p < .05; ** p < .01; *** p < .001.
Table 2.2

Meta-analytic standardized mean differences (fake - honest) for the effect of instructed faking on the domain and facets of psychopathy.

<table>
<thead>
<tr>
<th>Psychopathy</th>
<th>k</th>
<th>N</th>
<th>p</th>
<th>ES</th>
<th>95% CI</th>
<th>P</th>
<th>t²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>faking condition</td>
<td>honest condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total psychopathy</td>
<td>17</td>
<td>1325</td>
<td>1338</td>
<td>.002</td>
<td>-0.615</td>
<td>[.22,1.01]</td>
</tr>
<tr>
<td></td>
<td>Primary psychopathy</td>
<td>6</td>
<td>298</td>
<td>298</td>
<td>&lt;.001</td>
<td>-0.837</td>
<td>[.63,1.04]</td>
</tr>
<tr>
<td></td>
<td>Secondary psychopathy</td>
<td>6</td>
<td>298</td>
<td>298</td>
<td>.008</td>
<td>-0.580</td>
<td>[.15,1.01]</td>
</tr>
<tr>
<td></td>
<td>Total psychopathy</td>
<td>6</td>
<td>453</td>
<td>486</td>
<td>&lt;.001</td>
<td>1.806</td>
<td>[-2.17,-1.46]</td>
</tr>
<tr>
<td></td>
<td>Primary psychopathy</td>
<td>4</td>
<td>181</td>
<td>181</td>
<td>&lt;.001</td>
<td>0.895</td>
<td>[-1.28,-.51]</td>
</tr>
<tr>
<td></td>
<td>Secondary psychopathy</td>
<td>4</td>
<td>181</td>
<td>181</td>
<td>&lt;.001</td>
<td>1.076</td>
<td>[-1.52,-.63]</td>
</tr>
</tbody>
</table>

Note: n = cumulative sample size; k = number of independent studies; ES = summary effect size (Cohen’s d, calculated as fake – honest, such that positive values indicate an increase in scores under instructed faking, error bars signify variance within the sample of each study).
Effects of Instructed Faking on Psychopathy Facets

Hypothesis 1: Score changes under instructed faking. A subgroups analysis was performed to determine if these differences were a function of psychopathy at the facet level, i.e., primary or secondary psychopathy. For faking good instructions \((k = 6, N = 298)\), relative to honest responding, there was a significant score increase for both primary psychopathy \((d = -0.84, p < .001)\) and secondary psychopathy \((d = -0.58, p = .008)\). For faking bad instructions \((k = 4, N = 181)\), there was a significant score decrease for both primary psychopathy \((d = 0.90, p < .001)\) and secondary psychopathy \((d = 1.08, p < .001)\). These results support Hypothesis 1.

Hypothesis 2: Magnitude of ‘fake bad’ score changes is larger than ‘fake good’. Hypothesis 2 was tested by checking whether confidence intervals for ‘fake good’ versus ‘fake bad’ effects overlapped. Confidence intervals, as reported in Table 2.2, overlapped for both primary psychopathy and secondary psychopathy, such that differences in magnitude were not significant. For primary psychopathy, this difference was also not significant when a meta-regression was performed using the absolute value of the effect size as the criterion \((b = 0.058, p > .05)\). Similarly, secondary psychopathy was not significant when a meta-regression was performed using the absolute value of the effect size as the criterion \((b = 0.496, p > .05)\). Hypothesis 2 was not supported for primary or secondary psychopathy.

Publication Bias

We assessed the likelihood of publication bias by inspecting the funnel plots for fake good versus honest instructions (see Figure 2) and fake bad versus honest instructions (see Figure 3) of the relationship between observed effects and standard error for symmetry (Schwarzer et al., 2015). Egger’s test was also run by including standard error as a predictor in the subgroups analysis. Based on the funnel plots and a non-significant Egger’s test of
asymmetry (p = .18), risk of publication bias was judged to be low. Egger’s test is a commonly used assessment of publication bias (Schwarzer et al., 2015).

**Figure 2.2**

*Fake Good and Fake Bad Funnel Plot*

2.5 Discussion

This meta-analysis and systematic review shows that instructed faking has similar effect sizes for the dark triad compared to the Big Five personality domains. That is, instructions to ‘fake good’ result in lower psychopathy and Machiavellianism scores, whereas instructions to ‘fake bad’ result in higher psychopathy and Machiavellianism scores (in line with our first hypothesis). The single study addressing narcissism showed the surprising finding that ‘fake good’ instructions produced in *higher* narcissism scores (contrary to our first hypothesis). Both the systematic review of Machiavellianism and the meta-analysis of the psychopathy domain showed larger effects for ‘fake bad’ than ‘fake good’ instructions. This supported our second hypothesis and is consistent with prior meta-analytic findings for the Big Five traits (Viswesvaran & Ones, 1999). That is, our results show that dark triad personality tests are just as fakeable as Big Five personality tests, with particularly marked effects for ‘faking bad’. Although, if an individual’s baseline responses are already positively biased then it may be that individuals have less room to ‘fake good’ but more potential to ‘fake bad’. That is, the smaller ‘fake good’ effect size may just be an artefact caused by a ceiling effect. Publication bias was assessed to be low in this meta-analysis. As is typical in the area of instructed faking, effect sizes are typically large (*d*-scores of between 1 and 3).
Additionally, applied research is generally seeking non-significant effect sizes or smaller effects, to demonstrate an improvement in assessment practices. For these reasons, it is unlikely that these criticisms of publication bias apply to this area of research.

In the paragraphs below, we focus our discussion on two main issues that follow from our results. First, we consider the process of faking, and how this may apply to dark triad assessments. We consider the distinction between descriptive and evaluative item content, and how this distinction might be used to develop assessments of dark triad concepts. Second, we consider the meaning and implications of our results for the applied use of the dark triad psychometric tests. In contrast to the Big Five personality traits (where tests are mainly used for job selection), dark triad assessments (particularly psychopathy) are used for forensic applications, including risk assessment for sex offenders but also for ‘screening out’ those with psychologically deviant tendencies from law enforcement and armed services. Arguably, an error in releasing an offender into the community or admitting a psychopathic individual into law enforcement has more serious consequences than failing to select the best candidate for a graduate or managerial job.

**Descriptive versus Evaluative Item Content: Implications for Response Distortion**

Personality items have both descriptive content (i.e., the construct being measured) and evaluative content (i.e., the extent to which it sounds positive versus negative). Both Anglim et al. (2017) and Bäckström et al., (2009) suggest that faking involves an item level evaluation of social desirability. That is, people decide to adjust their score on each item based on how positive or negative-sounding the item content is. In the case of the dark triad, items are transparently negative (i.e., they sound like a bad thing to have according to social norms). Expert ratings of the social desirability of key dark triad rating scale items found the evaluative content of items were rated by experts as either largely socially desirable or undesirable, with very few items being rated as neutral (Walker et al., 2020).
It is not a new observation that items’ evaluative content may enhance socially desirable responses on those items. However, research has mainly focused on the Big-Five rather than the dark triad. Edwards (1957) found rating scales of personality inventories tend to correlate almost perfectly with expert ratings of an item’s social desirability. Individuals endorsing positively-valanced items while denying negatively-valanced items suggests the covariance between items, and even measures (for a review of the Machiavellianism/psychopathy measurement issues see Miller et al., 2017) is due to the shared evaluative content (i.e., positive or negative valance) of dark triad scales rather than descriptive content. In support of this idea, Bäckström (2007) found significantly smaller correlations among the big five traits after controlling for the evaluative content.

Further, when Bäckström et al., (2009) presented Big-Five items phrased neutrally, the covariance between items was substantially reduced providing a clearer factor structure of the Big-Five. Essentially, Bäckström et al. (2009) concluded that neutral items elicited less socially desirable responding than positively or negatively valanced evaluative items. One possible future research direction for dark triad assessment development would be to focus on neutralizing the evaluative content. This is both a concern for applied use (as positive evaluative content elicits faking) but also examining the true relationship of the dark triad traits with each other and with valued outcomes once the confounding influence of negative evaluative content is removed. However, there are some potential roadblocks to neutralising dark triad scale items. Items on dark triad measures are constructed in such a way to connect with the necessarily darker tendencies and attitudes of the constructs being assessed. As a result, it may be difficult to remove the evaluative content of the items while still maintaining the integrity of the descriptive content. The currently open question of whether dark triad scale items measure beyond the evaluative nature of the constructs should be explored.
There is evidence that standard dark triad assessments’ evaluative item content is more socially undesirable for psychopathy and Machiavellianism than narcissism (Walker et al., 2020). If faking is based on an item-level evaluation of social desirability, it would then follow that people should show more pronounced faking on psychopathy as compared to narcissism. In particular, some items representing grandiose narcissism (ego-centric self-confidence and belief in superiority) have positive evaluative content. In fact, expert ratings of the social desirability of narcissism items found items measuring grandiose narcissism were more socially desirable than undesirable (Walker et al., 2020). The study in our systematic review of narcissism (Young, 2018) included items such as ‘I am more capable than other people’ and ‘People see me as a natural leader’, which arguably sound highly desirable. Young’s (2018) scale items cover the domain of grandiose narcissism but do not include items measuring vulnerable narcissism. The inclusion of these and similar items in Young (2018) may explain why fake-good instructions resulted in higher narcissism scores.

This distinction between grandiose narcissism (which can have relatively positive evaluative item content) and vulnerable narcissism (which generally does not) points to the critical importance of considering dark triad facets as well as domains. Few of the studies we have reviewed differentiated the dark triad traits at the facet level. This may be a particularly important future consideration given substantial heterogeneity within narcissism (Miller et al., 2011) and psychopathy (Newman et al., 2005). For example, grandiose narcissism tends to be positively related to Extraversion and negatively related to Agreeableness and Neuroticism, while vulnerable narcissism is typically negatively related to Extraversion and positively related to Neuroticism (Paulhus & Williams, 2002). Similarly, there are differences between primary and secondary psychopathy, such that Gray’s behavioural inhibition system (BIS) is weakly associated with primary psychopathy and the behavioural activation system (BAS) is strongly associated with secondary psychopathy while only moderately associated
with primary psychopathy (Newman et al., 2005). Additionally, a recent systematic review found that while total psychopathy and total narcissism are not consistently related to emotional intelligence, grandiose narcissism and primary psychopathy tend to positively relate to emotional intelligence, while vulnerable narcissism and secondary psychopathy tend to be negatively related to emotional intelligence (Walker et al., 2020). The collapsing of facets into total scores on dark triad measures is common within the dark triad, and faking literature. However, representation of total scores without consideration of the facets hinders the development of a unified theoretical and empirical framework of the dark triad (Miller et al., 2011). A unified framework necessarily leads to improved measurement of these traits, and enhanced ability to accurately measure the extent to which the scales are fakeable.

**Implications of Faking for Applied Use of Dark Triad and Related Measures**

Our results demonstrate that the dark triad assessments are fakeable (i.e., people *can* fake on the tests to get a higher score). An important additional consideration is whether in practice, people really *do* fake when motivated to do so. Faking research on the Big Five indicates that people *do* fake, though not all people fake to the maximum possible extent all the time (Birkeland et al., 2006; Griffith & Converse, 2012; Viswesvaran & Ones, 1999). Moreover, people do not fake indiscriminately but rather tend to alter their responses to match their perception of desirable traits required for specific job types (Furnham, 1990). When people are asked to describe an ideal employee, they create a response profile tailored to closely match the attributes of the ideal (Martin et al., 2002), or typical (Levashina & Campion, 2006) employee. While some research has suggested there is little response distortion in real-world recruitment settings (Hough et al., 1990), other research reports job applicants tend to rate themselves more highly on big-five domains compared to job incumbents (Birkeland et al., 2006; Griffin et al., 2004). In fact, incumbent employees wishing to appear socially desirable to their employers may be motivated to represent
themselves as having higher levels of desirable traits (Higgins et al., 2007). Similar response patterns emerge when using the HEXACO. Individuals in job applicant settings represent themselves significantly more favourably than individuals in low-risk settings (Anglim et al., 2017). Taken together, there is evidence to suggest personality rating scales’ susceptibility to faking is capitalized on by applicants in a selection context – people exploit the fakeability of tests to gain higher scores for themselves and do so in a reasonably sophisticated way. Dark triad assessments are frequently used for forensic applications, where outcomes may be even higher stakes for the test-taker (e.g., psychopathy scores may partly determine incarceration). Given that people are likely to fake under such circumstances, the extent to which faking can alter scores is a critical consideration. We consider the implications of our results for several possible contexts below: screening for law enforcement employment, preventative detention proceedings for sex offenders, and selection contexts more generally.

**Law enforcement selection.** Over 90% of USA police agencies use psychometric assessment to screen out applicants with psychological deviance (Cochrane et al., 2003). The most commonly used instrument is the Minnesota Multiphasic Personality Assessment (MMPI), which focuses on ‘dark’ content that has considerable content overlap with the dark triad, particularly psychopathy (e.g., content scales include anger, cynicism, antisocial practices, psychotic thought processes [bizarre mentation] and low self-esteem). Current worldwide events highlight the global importance of fake-proof selection procedures to screen out law enforcement officers with psychopathic tendencies. While the MMPI includes several possible lie scales intended to detect different kinds of faking (fake-good, fake-bad, and feigning particular symptom clusters), accuracy rates for detecting faking vary, and there are concerns about how to balance the needs of sensitivity (rate of detecting true faking) versus specificity (rate of detecting honest responding). This is particularly vexed given the
possibility of litigation for unfair hiring practice should a classification error occur (e.g., an applicant denied employment due to untrue accusations of ‘faking good’).

**Preventative detention proceedings for sex offenders.** Psychopathy assessments are often used as a criterion for determining whether sex offenders should be subject to civil commitment (i.e., ‘sectioned’ or detained involuntarily). These evaluations frequently use the Psychopathy Checklist Revised (PCL; Boccaccini et al., 2017). Although these evaluations are clinician-ratings rather than self-ratings, the possibility of faking/distortion from motivated evaluators has been noted as a concern (Boccaccini et al., 2017). Specifically, Murrie et al., (2008) found that opposing legal teams had different PCL scores in the direction corresponding to their prosecution versus defense stance. These differences were larger and more consistent in direction than would be predicted by chance, suggesting some motivated response distortion on the part of the evaluators. The possibility that observer-ratings are also subject to response distortion has received little attention in the research literature yet is critically important for forensic applications as well as ‘referee-report’ style personality ratings for job selection. We suggest that that the faking in observer ratings of personality constitutes an important and understudied area for future research.

**The increasing use of ‘dark’ personality assessments for selection.** The distinction between the generally positive Big-Five and the generally negative dark triad is important for considering applications of personality testing. Using ‘darker’ personality measures may be important for identifying potentially deviant behavior in people who hold positions of power over others. This is important in selection contexts that have traditionally used the Big Five. For example, numerous senior managers across a range of organizations and industries have been the center of highly publicized misconduct investigations (Van Scotter & De Dea Roglio, 2018). Additionally, there is a wide literature on selection criteria and risk factors associated with misconduct for medical school (Munro et al., 2005; Bore et al., 2009; Yates
& James, 2010), teachers (Knoll, 2010; Walter, 2018), and therapists (Gray & Garrett, 2017) amongst others. The inclusion of dark triad measures for identifying potential risk factors such as impulsivity, exploitation of others, manipulation, poor management etc., may therefore provide a better selection process than using Big Five measures alone. However, before widespread use of these measures is recommended a clearer understanding of the extent to which the validity of these measures is undermined by faking is necessary.

**Conclusion**

Our review found that dark triad assessments were fakeable, with similar effect sizes to previous meta-analyses of Big Five measures. Following from this evidence, we argue for caution in interpreting dark triad assessments as the basis for high stakes decision making, definitely for self-ratings but also for observer-ratings from another source. We suggest that future research could focus on facet-level assessment of dark triad traits, manipulating or removing evaluative item content, and examining response distortion in observer-ratings as well as self-ratings.
References


Karpman, B. (1941). On the need of separating psychopathy into two distinct clinical types: the symptomatic and the idiopathic. *Journal of Criminal Psychopathology, 3*, 112-137


CHAPTER 3: Faking Good and Bad on Self-Reports versus Informant-Reports of Dark Triad Personality
3.1 Abstract

Research consistently demonstrates that people can fake on self-report personality tests. Informant-reports (where a knowledgeable informant rates a target’s personality) can be used as an alternative to self-ratings. As yet, no studies have investigated whether informants can also fake on personality inventories or what may motivate faking. Study 1 examines the effects of experimentally-induced faking on self- and informant-reports of the Dark Triad of personality and whether self/other-interested motivation may influence the extent to which people fake. In Study 1 ($N = 834$ undergraduates), participants completed Dark Triad measures in a 2x3 between-person design crossing format (self- versus informant-report) with instruction condition (answer honestly, fake good, or fake bad). Results indicated significant faking good and faking bad, on both self- and informant-reports. To determine whether this was influenced by the evaluative content of items on Dark Triad scales, Study 2 asked expert personality researchers ($N = 9$) to rate the social undesirability of the Dark Triad scale items used in study 1. Strong interrater reliability among the experts found the items to be highly socially undesirable. Studies 1 and 2 indicate that people endorse items (for self- and informant-reports) based on the items’ evaluative content in both faking and honest conditions. Implications for the use and interpretation of Dark Triad scales are discussed.

Keywords: dark triad, narcissism, Machiavellianism, psychopathy, evaluative content, instructed faking, informant-report, self-report
3.2 Introduction

Although self-report scales’ accuracy has often been criticized (Paulhus & Vazire, 2007), self-report rating scales are the most frequently used method to assess personality traits. One major criticism is that self-ratings allow the test-taker to engage in response distortion. That is, test-takers may attempt to make themselves look better than they really are (to “fake good”) or worse than they really are (to “fake bad”), resulting in inaccurate measurement (Edwards, 1957; Paulhus, 2002). Meta-analyses show that test-takers can distort their responses when instructed to do so, resulting in large to very large changes to mean scores on both Big Five and Dark Triad personality traits (Viswesvaran & Ones, 1999; Walker et al., 2020). One suggested solution to faking on self-report rating scales is to use informant-reports (Kim, et al., 2019; Lukoff, 2012; Vazire, 2006). Prior meta-analytic findings show satisfactory agreement between self- and informant-reported personality scores, emphasizing the benefits of using both self- and informant-reports to measure personality (Connelly & Ones, 2010; Kim et al., 2019). However, it is possible that informants might also engage in faking. As yet, there is has been little to no empirical examination of informant-faking. That is, no study has investigated the extent to which informants can “fake good” or “fake bad” on behalf of someone else on either Big Five or Dark Triad rating scales.

In contrast to the Big Five personality traits (where tests are mainly used for job selection), Dark Triad assessments (particularly psychopathy) are used for forensic applications, including risk assessment for parole hearings but also for ‘screening out’ those with psychologically deviant tendencies from law enforcement and armed services. Arguably, an error in releasing an offender into the community or admitting a psychopathic individual into law enforcement, or even into high ranking governmental positions has more serious consequences than failing to select the best candidate for a graduate or managerial job. For
this reason, we focus on faking on the Dark Triad and examine the extent to which people distort their responses on Dark Triad domains and facets when instructed to fake for themselves or someone else.

**Instructed Faking**

Despite the validity of self-reported information (Holden & Passey, 2010), there has been substantial debate about the accuracy of self-report scales, which represents an ongoing challenge for researchers (Hogan & Foster, 2016). These concerns typically relate to the impact of response sets, biases, and styles on the validity of self-reported non-cognitive data. Not all test-takers respond accurately. Some may distort responses unconsciously, while others may intentionally distort their responses when motivated to do so (Edwards, 1957; Paulhus & Vazire, 2007). Ego-protective or self-serving biases tend to arise when an individual’s sense of self is threatened. An individual may endorse items to enhance or protect their self-perceptions (Rodman et al., 2017).

Instructed faking paradigms are a standard method used to examine the extent to which individuals *can* distort their responses in a socially desirable way on personality measures. Individuals may intentionally present themselves more favorably than they really are (*faking good*) or less favorably than they really are (*faking bad*, also referred to as malingering; Arthur et al., 2010; Donaldson & Grant-Vallone, 2002; Furnham, 1990; Rogers et al., 2003). For example, a job applicant may try to enhance their positive qualities to look like a good fit for the job (Birkeland et al., 2006). In contrast, test-takers may *fake-bad* to make themselves look worse than they really are. For example, in a forensic setting, an offender may inflate their negative qualities when being assessed for competency to stand trial (Mills & Kroner, 2005).

Typically, instructed faking research has focused on the Big-Five model of personality in which extensive literature has demonstrated that people *can* fake on personality
measures (Viswesvaran & Ones, 1999). However, a recent meta-analysis and systematic review demonstrated people could also fake on Dark Triad measures. Cohen’s $d$ effect sizes on psychopathy measures ranged from -0.61 for faking-good and 1.81 for faking-bad (Walker et al., 2020).

**Self- and Informant-reports**

Despite enduring concerns people may distort their responses, self-report rating scales are the most commonly-used method for measuring personality traits (Griffith et al., 2007). Self-report rating scales represent the most direct route to obtaining information about a person, but people may not always respond accurately. This lack of accuracy can be due to an unrealistically optimistic view of the self (i.e., self-deceptive enhancement or denial; Paulhus, 1990) or a deliberate attempt to create an overly-positive false impression (i.e., impression management; Paulhus, 1990). The accuracy of self-report rating scales relies on the test-taker’s truthful responding and the assumption of accurate self-perception (Klonsky et al., 2002; Paulhus & Vazire, 2007).

An alternate method of obtaining information about a person is through informant-reports. Informant-reports are assessments in which another person (i.e., work colleague, friend, parent) rates target (Vazire, 2010). How well an informant knows the target and how much the informant likes the target can impact the informant’s reporting (Hollander, 1956; Leising et al., 2010). Indeed, early users of informant-reports raised concerns about potential “friendship effects”, suggesting informants with a close relationship to the target may respond favorably about them (Hollander, 1956). Such “friendship effect” concerns are often dismissed with the suggestion that informants are not motivated to distort their responses as the consequences of their responses do not directly impact them. However, Leising et al. (2010) demonstrated the degree to which an informant likes the target inevitably impacts their rating of the target. Recent findings tend to support this view with Beckmann et al.
(2020) reporting that informant ratings of Big Five are more favourable than self-ratings for all five traits (i.e., informants ratings were higher than self-ratings for extraversion, agreeableness, conscientiousness, and openness but lower for neuroticism). Beckmann et al.’s effects were larger for non-work informants (romantic partners, friends, and/or family members) than for work informants (direct reports, supervisors and/or work peers), which would support a “friendship effect”.

In contrast to Beckmann et al. (2020), Mount et al. (1994) found self-ratings were more favourable than informant ratings across all Big Five personality traits. Like Leising et al. (2010), Mount et al. found informants who were well acquainted with the target tended to rate the target more favourably than informants with lower acquaintance of the target, though never as favourably as the target rated themselves. Similarly, Clifton et al. (2005), Miller et al. (2015) and Sleep et al. (2017) reported self-reported ratings of narcissism were more favorable (i.e., lower narcissism) than informant ratings. Psychopathic traits were also rated more favorably (i.e., lower psychopathy) when self-reported compared to when rated by an informant (Miller et al., 2011). Taken together, these findings indicate that self-ratings are likely to be generally more favourable than informant-ratings on all dark triad traits. That is, informant-ratings will be generally higher for narcissism, Machiavellianism, and psychopathy domains and facets.

**Self-Other Interest**

It is widely accepted that people behave according to their own self-interests, assuming others also act the same (Miller & Ratner, 2001). However, people have also shown they are willing to act in the interests of other people (Gerbasi & Prentice, 2013). The Self/Other-Interest Inventory (SOII; Gerbasi & Prentice, 2013) was developed to assess the separate constructs of self-interest and other-interest. Self-interest refers to a deliberative process in the pursuit of personal gain. Other-interest is related to self-interest in that it is a
deliberative process pursuing gains for other people. Gerbasi and Prentice’s scale empirically measures the early eighteenth-century theory that people are willing to regard others’ interests when doing so directly or indirectly benefits themselves in some way. Obtaining information about an individual’s self-interest and other-interest motivations may provide a valuable insight into why people fake for themselves, and whether faking on behalf of another person is due to an internal motivation to act in another persons’s interest.

The Dark Triad

The Dark Triad represents a set of three socially aversive traits comprising narcissism, Machiavellianism, and psychopathy (Paulhus & Williams, 2002). Although conceptually distinct from each other, there are empirically overlapping characteristics such as ethical, moral, and socially deviant behavior as well as sharing an interpersonally exploitative demeanor with goal-focused manipulation of others’ emotions to get what they want. Given the Dark Triad predicts a range of socially aversive behavior in the workplace, for example, (for a review see Spain et al., 2014) it is unsurprising interest in Dark Triad assessment has gained popularity for its use of screening out people possessing maladaptive traits (Spain et al., 2014).

As a personality trait, narcissism is measured on a spectrum from mild to extreme (Miller & Campbell, 2008; Raskin & Hall, 1979) and comprises two facets: grandiose and vulnerable narcissism. Grandiose narcissism is characterized by grandiosity, self-confidence, and exploitation of others with a tendency to rely on self-internal validation (Dickinson & Pincus, 2003; Pincus et al., 2009; Raskin & Hall, 1979; Zhang et al., 2017). When threatened, individuals with high levels of grandiose narcissism tend to blame and devalue others while refusing to acknowledge their own weaknesses. Vulnerable narcissism is characterized by grandiose fantasies with oscillations between self-love and self-loathing typified by a fragile sense of self (Pincus et al., 2009; Raskin & Hall, 1979; Wink, 1991). Individuals with high
levels of vulnerable narcissism tend to rely on external validation for self-esteem maintenance with a hypersensitivity toward negative feedback, inevitably leading to defensiveness, resentment, and aggressive outbursts toward others when their sense of self is threatened (Wink, 1991). The multifaceted nature of narcissism has also been found to differentially relate to other constructs such as emotional intelligence such that grandiose narcissism tends to relate positively to emotional intelligence while vulnerable narcissism relates negatively (Walker et al., 2021). It is becoming much more common for research to examine the distinction between the two facets as it becomes more apparent that a total narcissism score is masking a much more nuanced understanding of the underlying mechanisms of the construct.

Machiavellianism is most commonly measured as a unidimensional personality construct derived from the philosophical writings of Niccolo Machiavelli, a political advisor to the Medici family in 16th century Firenze (Christie & Geis, 1970). Machiavellianism is characterized by goal-focused manipulative and callous social interactions (Christie & Geis, 1970). Individuals high in Machiavellianism tend to be regarded as capable of long-term strategic planning, delaying gratification for better rewards in the future, having questionable morals, and a cold, cynical world-view (Christie & Geis 1970; Furnham et al., 2013).

The last of the Dark Triad traits, psychopathy, is broadly understood as characterized by interpersonal, affective, and behavioral characteristics, including superficial charm, pathological lying, and lack of empathy, conscience, and remorse (Cleckley, 1951; Hare, 2003). Though often confused with antisocial personality disorder, sub-clinical psychopathy exists on a continuum in the wider population (Berg et al., 2013). Like narcissism, psychopathy comprises two related but distinct factors differing in their etiology and symptomology (Karpman, 1941). Hare (2003) expanded on Karpman’s previous work developing a model of psychopathy distinguishing primary and secondary psychopathy. Both
facets are typified by indifference to their own and others’ emotions, each underpinned by an antagonistic interpersonal style (Miller & Lynam 2012). Primary psychopathy is characterized by a lack of guilt and remorse, callousness, interpersonal manipulation, and having superficial affect (Casey et al., 2013; Hare, 2003). Additionally, primary psychopathy is associated with deficits in affective empathy (Wai & Tiliopoulos, 2012), lower levels of fear (Patrick et al., 1993), and neurotic anxiety (Newman & Brinkley, 1997). In contrast, secondary psychopathy is associated with higher neurotic anxiety levels alongside higher levels of antisocial behaviours such as aggressiveness and impulsivity (Levenson et al., 1995). Additionally, high secondary psychopathy levels are associated with guilt and fear responses not typically observed in individuals with high levels of primary psychopathy (Lykken, 1995; Wallace et al., 2009). Like narcissism, distinguishing between primary and secondary psychopathy has important implications. For example, primary psychopathy typically relates positively to emotional intelligence, while secondary psychopathy has a strong negative association with emotional intelligence (Walker et al., 2021). Further, considering the aetiology differs between primary and secondary psychopathy, it is likely the underlying motivations to respond untruthfully to a questionnaire may also be different, as too may be the extent to which people will fake on one facet compared to the other. For this reason, measuring the multi-faceted nature of psychopathy rather than just presenting a total score is relevant to this research.

The general aim of our two studies are to examine the extent to which self and informant-ratings of dark triad traits and facets are susceptible to response distortion under simulated ‘high stakes’ conditions of obtaining or avoiding employment. We consider the role of several possible factors in determining the extent of faking on the dark triad: the target (oneself versus a friend), the motivational condition (wanting a desirable job versus wanting to avoid an undesirable job), the trait (which dark triad domain or facet), individual
differences in *self-interest and other interest*, and the *social desirability of the constituent items* (as determined through expert ratings).

### 3.3 Study 1: Instructed Faking on Self- and Informant-reports

Study 1 investigates the extent to which people fake for themselves and others on the Dark Triad measures. Furthermore, to determine whether the extent to which they fake (for themselves or others) is influenced by the test-takers self- or other-interested attitudes. Our expectations regarding the effects of faking on self-report scales are primarily based on prior meta-analytic findings of the Dark Triad. In Walker et al.’s (2020) meta-analysis and systematic review of instructed faking on the Dark Triad, there were two broad trends. First, there were large effects of *faking-good* on all three Dark Triad domains, though not all in the expected direction - the single study examining narcissism found that scores increased rather than decreasing during *fake-good* instructions. Second, the effects were much larger for *fake-bad* than *fake-good* studies. Our expectations regarding the effects of faking on informant-report scales are largely based on prior findings demonstrating that informants inflate their target's positive characteristics (Leising et al., 2010). As such, we expect informants *can* fake informant-reports on Dark Triad measures when instructed to do so. Finally, as high self-interest levels predict people will behave in a way to benefit themselves, people with high self-interest levels will fake more for themselves than for others. (Gerbasi & Prentice, 2013). In contrast, when people are other-interested, that is, high other-interest predicts people will behave in a way to benefit others, they will fake more for others than themselves.

**Hypotheses**

**Hypothesis 1:** Individuals will fake on all measures of the Dark Triad. There will be a significant mean difference on all Dark Triad scores for ‘honest’ as compared to ‘fake good’ and fake bad’ conditions. We expect all Dark Triad scores to be higher in the ‘fake
bad’ condition (Hypothesis 1a) and lower in the ‘fake good’ condition (Hypothesis 1b) as compared to the ‘answer honestly’ condition.

**Hypothesis 2**: Individuals will give better scores to themselves than others. That is, all Dark Triad scores will be significantly lower for self-reports than informant-reports.

**Hypothesis 3**: There will be an interaction between instruction condition and ratings target, such that people will fake more for others than for themselves. That is, differences between the “fake good” and “answer honestly” conditions will be larger for informant- than self-reports (indicating more “faking good” for others than themselves; Hypothesis 3a) and the difference between “fake bad” and “answer honestly” will similarly be of larger for informant- than self-reports (Hypothesis 3b).

**Hypothesis 4**: Self/other-interest motivations will moderate the extent to which people fake for themselves compared to others. That is, people with high self-interest will show greater faking for themselves (as compared to others) whereas people with high other-interest will show greater faking for others (as compared to themselves).

### 3.3.1 Method

**Participants**

First year undergraduate psychology students (N = 834) at an Australian university took part in the study during an in-class activity. Included were 577 females (256 male) and 1 non-binary, with a mean age of 27.46 (SD = 3.07) ranging from 17 to 55 years of age. Participants were tested in groups of no more than 24 and participated as part of a tutorial lesson. There was an initial sample size of 927 participants. In accordance with our preregistered study plan (available at [http://aspredicted.org/blind.php?x=ke3wc3](http://aspredicted.org/blind.php?x=ke3wc3)), 56 participants were excluded for selecting “not well” or “not well at all” from the demographic question “how well do you speak English”. 34 participants were excluded for failing both
data check items. Three participants were removed for providing responses in free-text fields that did not answer the question, or did not make sense.

**Materials**

*Pathological Narcissism Inventory- Brief* (PNI-B; Schoenleber et al., 2015) is a 28-item scale rated on a 6-point Likert scale ranging from “not at all like me” to “very much like me”. These scales are aggregated into composites to represent grandiose narcissism (e.g., “I can usually talk my way out of anything”) and vulnerable narcissism (e.g., “I often hide my needs for fear that others will see me as needy and desperate”). The PNI-B also provides a total narcissism score.

*Levenson’s Self-Report Psychopathy Scale* (LRSP; Levenson et al., 1995) is a 26-item scale assessing psychopathic attitudes measured on a 4-point rating scale ranging from “disagree strongly” to “agree strongly”. There are two factors produced from this measure. (1) Primary psychopathy (e.g., “I enjoy manipulating other people’s feelings”); and (2) secondary psychopathy (e.g., “I am often bored”). Additionally, the LRSP provides a total psychopathy score.

*The Five Factor Machiavellianism Inventory* (FFMI; Collison et al., 2018) consists of 52-items answered on a 5-point Likert scale (disagree strongly to agree strongly) measuring sub-clinical Machiavellianism. This scales measures three facets of Machiavellianism including Antagonism (e.g., “Humility is overrated”), Agency (e.g., “I would rather be known as “practical” than “kind”), and Planfulness (“I like to map out my projects before I begin”).

*The Dirty Dozen* (Jonason & Webster, 2010) is a 12-item measure measuring the Dark Triad traits on a 5-point Likert scale (not well to extremely well). There are four narcissism items (e.g., “I tend to want others to admire me”), four Machiavellianism items (e.g., “I tend
to manipulate others to get my way”) and four psychopathy items (e.g., “I tend to lack remorse”). Data obtained from the dirty dozen was not included in the following analyses.

*Self-Other Interest Inventory* (SOII; Gerbasi & Prentice, 2013) consists of 18-items measured on a 7-point Likert scale ranging from “very much like me” to “not like me at all”. This scale measures two separate constructs (a) self-interested motivation (e.g., “I am always looking for ways to get ahead”), and (b) other-interested motivation (e.g., “I want to help people I know do well”).

**Procedure**

All participants completed a demographic questionnaire and the self/other-interest questionnaire under normal instructions. This study uses a 2 x 3 between-person design crossing rating type (self-report versus informant-report) with instructions (respond honest, fake good, or fake bad). Participants were randomly assigned to one of six conditions of this design using Qualtrics randomiser function. In the “informant-report” conditions, participants were asked to think of a peer of the same sex and age (not a romantic partner).

An example of the instructions for self-report fake good:

On the next screens, you will complete a personality rating scale and a situational test of personality. As you complete these tests, we want you to pretend that you are a job applicant for a job that you really want. Take a minute to think about a job that you think would be great. Imagine that the interviewer tells you that before she can hire you, you need to complete a couple of questionnaires to see if you really are a good fit for the company. *If you score well, you will get the job*. If you do not score well, you will not get the job. So, as you complete the next two questionnaires, we want you to respond to each statement by *describing yourself with the goal of making the company believe that you will make a good employee.*

An example of the instructions for informant-report fake bad:

*Think of a friend you have (who is not your partner) who is about the same age and gender as yourself.*

What was the gender of your friend?  MALE  FEMALE  OTHER

What is the age of your friend? ______ years
On the next screens, you will complete a rating scale and a situational test that ask about your friend’s personality. As you complete these tests, we want you to pretend that your friend is a job applicant for a job that they really do NOT want. Take a minute to think about a job that your friend would think was terrible. Imagine that the interviewer tells your that before the company can hire them, they need to ask a referee to complete a couple of questionnaires to see if your friend will be a good fit for the company. If your friend scores well, your friend will get the job. If your friend does not score well, your friend will not get the job. So, as you complete the next two questionnaires, we want you to respond to each statement by describing your friend with the goal of making the company believe that your friend will NOT make a good employee.

Following the instruction screen, participants were asked “what did the instructions ask you to do?” and participants had to select a response from several options (e.g., “Rate myself honestly”, “Complete some intelligence tests”). If they did not answer correctly, the instructions were displayed a second time. There were four manipulation checks in place to ensure attention to the task and instructions (See supplementary materials). All protocols were approved by the Human Research Ethics Committee of the first author’s institution. This study was pre-registered, and a copy of this pre-registration can be found at this link “https://aspredicted.org/blind.php?x=ke3wc3”.

Analysis

As stated in our pre-registration, Hypotheses 1 to 3 were tested using 3x2 ANOVAs. We used contrast coding as follows, for the 6 conditions informant/fake-bad, informant/fake-good, informant/honest, self/fake-bad, self/fake-good, self/honest: a) self-report versus informant-report (.33, .33, .33, -.33, -.33, -.33); b) fake-bad versus honest (0.5, 0 -0.5, 0.5, 0 -0.5), c) fake-good versus honest (0 0.5, -0.5, 0 0.5, -0.5); d) the interaction of faking bad with measurement type (i.e., contrast a * contrast b) and e) the interaction of faking good with

---

1 Note that we pre-registered dummy-coding (rather than contrast coding) for comparison of conditions (honest versus fake-good versus fake-bad). Contrast coding is therefore a minor departure from the pre-registration (but we feel that it is clearer). We did not conduct the dummy-coded analysis.
measurement type (i.e., contrast a * contrast c). A separate ANOVA was conducted for each of the Dark Triad domains and facets. We evaluated the effect size with respect to $\eta^2_P$, with values of .01, .06, and .14 considered as “small”, “medium” and “large” respectively (Cohen, 1988).

As stated in our pre-registration, hypothesis 4 was tested by running the above ANOVA as a regression (using the contrast coding specified above) and adding either: a) other-interest (mean-centred) and its interaction with the five contrast-coded variables; or b) self-interest (mean-centred) and its interaction with the five contrast-coded variables. We ran separate regression analyses for each of the Dark Triad domains and facets for both other-interest and for self-interest (i.e., 14 separate regressions, each with 11 predictors).

### 3.3.2 Results

**Reliability and Descriptive Statistics**

Table 3.1 presents the reliability, descriptive statistics, and mean differences under honest, fake-good and fake-bad instructions for self-reported and informant-reported Dark Triad trait and facet level scores. Internal consistency reliability for self-reported Dark Triad traits and facets was good across honest, fake good and fake bad conditions with the exception of the reliability for total psychopathy in the fake bad condition ($\alpha = .57$). Likewise, the internal consistency reliability for informant-reported Dark Triad traits and facets was also good with the exception of the total psychopathy honest condition ($\alpha = .67$), and fake bad condition ($\alpha = .64$). The reliability coefficients reported in this study are consistent with prior research for each of the Dark Triad measures (Kuckelhaus et al., 2020; Schoenleber et al., 2015; Selbom, 2011). The Cohen’s $d$ effect sizes for self-reported faking ranged from small to very large for faking good ($d = -1.22$ to 1.42), and very small to extremely large for faking bad ($d = -0.56$ to 3.58). Effect sizes for informant-reported faking
also ranged from small to very large for faking good ($d = -1.35$ to $0.62$), and large to extremely large for faking bad ($d = -0.55$ to $3.13$).
Table 3.1
Reliability and descriptive statistics for each dark triad domain and facet under answer honestly, fake good, and fake bad instruction conditions for self-report and informant-report (Cohen’s $d$ compares faking conditions to answer honestly)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Honest ($N = 281$)</th>
<th>Fake Good ($N = 281$)</th>
<th>Fake Bad ($N = 272$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td><strong>Self-report</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total narcissism</td>
<td>83.40</td>
<td>17.00</td>
<td>.89</td>
</tr>
<tr>
<td>Grandiose narcissism</td>
<td>39.10</td>
<td>7.38</td>
<td>.77</td>
</tr>
<tr>
<td>Vulnerable narcissism</td>
<td>44.20</td>
<td>11.50</td>
<td>.86</td>
</tr>
<tr>
<td>Machiavellianism</td>
<td>154.00</td>
<td>20.40</td>
<td>.85</td>
</tr>
<tr>
<td>Total psychopathy</td>
<td>72.5</td>
<td>11.30</td>
<td>.72</td>
</tr>
<tr>
<td>Primary psychopathy</td>
<td>34.50</td>
<td>8.72</td>
<td>.89</td>
</tr>
<tr>
<td>Secondary psychopathy</td>
<td>24.00</td>
<td>4.44</td>
<td>.89</td>
</tr>
<tr>
<td><strong>Informant-Report</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total narcissism</td>
<td>79.60</td>
<td>13.80</td>
<td>.83</td>
</tr>
<tr>
<td>Grandiose narcissism</td>
<td>36.40</td>
<td>7.13</td>
<td>.75</td>
</tr>
<tr>
<td>Vulnerable narcissism</td>
<td>43.20</td>
<td>9.59</td>
<td>.80</td>
</tr>
<tr>
<td>Machiavellianism</td>
<td>161.00</td>
<td>21.10</td>
<td>.85</td>
</tr>
<tr>
<td>Total psychopathy</td>
<td>73.00</td>
<td>11.30</td>
<td>.67</td>
</tr>
<tr>
<td>Primary psychopathy</td>
<td>36.10</td>
<td>8.63</td>
<td>.89</td>
</tr>
<tr>
<td>Secondary psychopathy</td>
<td>23.00</td>
<td>4.65</td>
<td>.82</td>
</tr>
</tbody>
</table>

*Note. Cohen’s $d$ compares the standardized mean difference for the honest condition to ‘fake good’ and ‘fake bad’ conditions. The $t$ statistic was derived from an independent samples t-test. Narcissism is measured by the Pathological Narcissism Inventory, Machiavellianism by the Five Factor Machiavellianism Scale, and psychopathy by the Levenson’s Self-report Psychopathy Scale (Levenson et al., 1995)
* $p < .05$, ** $p < .01$ *** $p < .00$
Hypothesis testing

The results of the ANOVAs (testing Hypotheses 1 to 3) are presented in Table 3.2. The results of the regressions (testing Hypothesis 4) are presented in Table 3.3.

**Hypothesis 1a: Dark Triad scores will be lower for “Fake Good” instructions.**

Table 3.2 shows that in comparison to the “answer honestly” condition, participants in the “fake good” condition showed significantly lower scores on two of the three Dark Triad domains (narcissism and psychopathy) and three of the four Dark Triad facets (vulnerable narcissism, primary psychology, and secondary psychopathy). The effect sizes on self- and informant-reports were: a) large for total narcissism; secondary psychopathy, and total psychopathy; (b) moderate for primary psychopathy; and (c) very small for vulnerable narcissism. These results are as in the hypothesized direction. However, scores did not significantly differ for honest versus fake good conditions for grandiose narcissism. For Machiavellianism, scores were significantly higher in the “fake good” as compared to the “answer honestly” conditions (the opposite direction to hypotheses), with a large effect size. We therefore have mixed support for Hypothesis 1a.

**Hypothesis 1b: Dark Triad scores will be higher for “Fake Bad” instructions.**

Table 3.2 shows that in comparison to the “answer honestly” condition, participants in the “fake bad” condition showed significantly lower scores on two of the three Dark Triad domains (narcissism and psychopathy) and three of the four Dark Triad facets (vulnerable narcissism, primary psychology, and secondary psychopathy). The effect size was: a) very large for vulnerable, primary psychopathy, secondary psychopathy, and total psychopathy. These results are as hypothesized. However, scores did not significantly differ for honest versus fake bad conditions for grandiose narcissism. For Machiavellianism, scores were significantly lower in the “fake bad” as compared to the “answer honestly” conditions (the
opposite direction to hypotheses), with a large effect size. We therefore also have mixed support for Hypothesis 1b.

**Hypothesis 2: Individuals will give better scores to themselves than others**

As shown in Table 3.2, contrast 3 tests the main effect of whether self-ratings are higher (i.e., “worse”) than informant-ratings. For Narcissism, there were no significant differences between self- and informant-ratings. For Machiavellianism, the effect was in the opposite direction to hypotheses. Self-ratings were significantly lower than informant ratings with a small effect size. For psychopathy, the total score and both facets showed significant effects in the hypothesized direction. For total psychopathy, self-reports were higher than informant reports, and this difference was significant but small. For primary psychopathy, self-reports were higher than informant reports, and this different was significant but small. For secondary psychopathy, self-reports were higher than informant reports, and this different was significant but small.

In summary, Hypothesis 2 was supported for psychopathy, but not for Narcissism or Machiavellianism. However, Machiavellianism results for Hypothesis 1 indicate that Machiavellianism may be considered a socially desirable trait (at least as measured by the FFFM), as test-takers increased scores under “fake good” instructions, but decreased scores under “fake bad” instructions. Therefore, results for Machiavellianism are consistent with participants generally giving “better” (more socially desirable) scores to others as compared to themselves.

**Hypothesis 3: People will fake more for others than for themselves.**

**Differences in “faking good” for informant versus self-ratings.** The interaction effects for “fake good” instructions with rater-type were significant for Machiavellianism, total psychopathy, and primary psychopathy (but not for narcissism or either of its facets). For Machiavellianism, instructions to “fake good” resulted in a very large score increase to
self-ratings and a moderate to large score increase for informant ratings. There was a small, but significant difference between self- and informant-ratings such that there is greater faking on self-ratings than informant-ratings (i.e., both the direction of faking and the difference between self-ratings and informant-ratings were in opposite directions to hypotheses). For total psychopathy there was a moderate difference between “fake good” and “answer honestly” for self-ratings and a very large difference for informant-ratings. This difference between self- and informant-ratings was small but significant, indicating greater faking for informant-ratings than self-ratings. For primary psychopathy there was a small difference between “fake good” and “answer honestly” for self-ratings and a large difference for informant-ratings. This difference between self- and informant-ratings was small, but significant, indicating greater faking for informant-ratings than self-ratings. Hypothesis 3a was thus supported only for total and primary psychopathy, and not for the other Dark Triad traits.

**Differences in “faking bad” for informant versus self-ratings.** The interaction effects of “fake bad” instructions with self- versus informant-ratings were significant only for grandiose narcissism and primary psychopathy. For grandiose narcissism, the difference between “fake bad” and “answer honestly” was very small for self-reports but small to moderate for informant-reports. The difference between self- and informant-ratings was significant with a small effect size, indicating that people were faking more for others than themselves. For primary psychopathy, there was an extremely large difference between “faking bad” and “answer honestly” for both self- and informant-ratings. The difference between self- and informant-ratings was significant with a small effect size, indicating that people were faking more for themselves than for others. While results were mixed, overall there is no support for the idea that people “fake bad” more for others than themselves.
Table 3.2
Results of the 3x2 between-subjects ANOVA testing instruction condition by target (self-report versus informant-report) and interactions.

<table>
<thead>
<tr>
<th>Narcissism</th>
<th>Total Narcissism</th>
<th>Grandiose Narcissism</th>
<th>Vulnerable Narcissism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast</td>
<td>Contrast</td>
<td>η²</td>
<td>Contrast</td>
</tr>
<tr>
<td>ψ1: honest vs fake good</td>
<td>-13.81</td>
<td>107.93**</td>
<td>-1.07</td>
</tr>
<tr>
<td>ψ2: honest vs fake bad</td>
<td>21.30</td>
<td>252.90**</td>
<td>1.30</td>
</tr>
<tr>
<td>ψ3: self vs informant</td>
<td>-1.63</td>
<td>2.24</td>
<td>-1.14</td>
</tr>
<tr>
<td>ψ4: interaction 1 (ψ1 x ψ3)</td>
<td>-1.37</td>
<td>0.27</td>
<td>-0.56</td>
</tr>
<tr>
<td>ψ5: interaction 2 (ψ2 x ψ3)</td>
<td>-5.00</td>
<td>3.49</td>
<td>-5.24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mach</th>
<th>Total Machiavellianism</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast</td>
<td>F</td>
<td>η²</td>
</tr>
<tr>
<td>ψ1: honest vs fake good</td>
<td>17.48</td>
<td>129.58***</td>
</tr>
<tr>
<td>ψ2: honest vs fake bad</td>
<td>-11.35</td>
<td>53.79***</td>
</tr>
<tr>
<td>ψ3: self vs informant</td>
<td>3.09</td>
<td>6.00**</td>
</tr>
<tr>
<td>ψ4: interaction 1 (ψ1 x ψ3)</td>
<td>13.32</td>
<td>18.82***</td>
</tr>
<tr>
<td>ψ5: interaction 2 (ψ2 x ψ3)</td>
<td>-1.87</td>
<td>0.36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Psychopathy</th>
<th>Total Psychopathy</th>
<th>Primary Psychopathy</th>
<th>Secondary Psychopathy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast</td>
<td>F</td>
<td>η²</td>
<td>Contrast</td>
</tr>
<tr>
<td>ψ1: honest vs fake good</td>
<td>-11.72</td>
<td>126.84**</td>
<td>-5.55</td>
</tr>
<tr>
<td>ψ2: honest vs fake bad</td>
<td>40.10</td>
<td>1462.54**</td>
<td>25.83</td>
</tr>
<tr>
<td>ψ3: self vs informant</td>
<td>-2.35</td>
<td>7.59**</td>
<td>-1.39</td>
</tr>
<tr>
<td>ψ4: interaction 1 (ψ1 x ψ3)</td>
<td>4.83</td>
<td>5.40*</td>
<td>5.04</td>
</tr>
<tr>
<td>ψ5: interaction 2 (ψ2 x ψ3)</td>
<td>3.88</td>
<td>3.43</td>
<td>3.97</td>
</tr>
</tbody>
</table>

Note. The contrasts for conditions Informant/Fake-Bad, Informant/Fake-Good, Informant/Honest, Self/Fake-Bad, Self/Fake-Good, Self/Honest were coded (0, 0.5, -0.5, 0, 0.5, -0.5), (0.5, 0, -0.5, 0.5, 0, -0.5), (0.333, 0.333, 0.333, -0.333, -0.333, -0.333), (0, 0.1665, -0.1665, 0, -0.1665, 0.1665), and (0.1665, 0, -0.1665, -0.1665, 0, 0.1665) to test contrasts 1 to 5.

* p < .05, ** p < .01
Hypothesis 4: Moderation of Self and Other-interest motivation.

Self-interest as a moderator of faking. As shown in Table 3.3, after controlling for instructions (honest, fake good and fake bad), target (self or other), and the target/instructions interactions, self-interest was significantly positively related to all Dark Triad traits except for secondary psychopathy. This indicates that people with high self-interest scores also have higher Dark Triad scores.

For most Dark Triad scores (all but vulnerable narcissism and secondary psychopathy) the relationship between self-interest and Dark Triad scores differed for “answer honestly” versus “fake bad”. Specifically, self-interest had a stronger association with Dark Triad scores in the “honest” condition compared to the “fake bad” condition.

Other interactions only held for a small number of Dark Triad traits. For total narcissism and grandiose narcissism, there was a significant interaction of self-interest with self- versus informant-ratings, such that narcissism showed a stronger positive association with self-interest for self-ratings compared to informant-ratings. That is, self-ratings of narcissism were positively correlated with self-interest whereas informant ratings were not. For grandiose narcissism only, there was a significant interaction of self-interest with honest versus fake good conditions. Specifically, the grandiose narcissism/self-interest association was stronger for answer honestly than fake good. That is, Hypothesis 4a is supported for total and grandiose narcissism only.
Table 3.3.

Standardized beta-weights for regressions predicting dark triad traits from contrast-coded conditions (self/informant, and fake-good/fake-bad/honest), adding self-interest, (plus self-interest interactions) or other interest (and other-interest interactions)

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Narcissism</th>
<th>Machiavellianism</th>
<th>Total</th>
<th>Psychopathy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grandiose</td>
<td>Vulnerable</td>
<td></td>
<td>Primary</td>
<td>Secondary</td>
</tr>
<tr>
<td><strong>Self-interest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\psi_1): honest vs fake good</td>
<td>- .62**</td>
<td>-.10**</td>
<td>-.73**</td>
<td>.57**</td>
<td>- .69**</td>
</tr>
<tr>
<td>(\psi_2): honest vs fake bad</td>
<td>.71**</td>
<td>.11**</td>
<td>.84**</td>
<td>-.50**</td>
<td>.99**</td>
</tr>
<tr>
<td>(\psi_3): self vs informant</td>
<td>-.03</td>
<td>-.06</td>
<td>-.01</td>
<td>.07*</td>
<td>-.05**</td>
</tr>
<tr>
<td>(\psi_4): good x self/informant ((\psi_1 \times \psi_3))</td>
<td>.00</td>
<td>-.09*</td>
<td>.04</td>
<td>-.17**</td>
<td>-.03</td>
</tr>
<tr>
<td>(\psi_5): bad x self/informant ((\psi_2 \times \psi_3))</td>
<td>.05</td>
<td>.17**</td>
<td>-.02</td>
<td>.10**</td>
<td>-.02</td>
</tr>
<tr>
<td><strong>Self-interest by (\psi_1)</strong></td>
<td>.12**</td>
<td>.12**</td>
<td>.08**</td>
<td>.06*</td>
<td>.04*</td>
</tr>
<tr>
<td><strong>Self-interest by (\psi_2)</strong></td>
<td>.05</td>
<td>.08*</td>
<td>.02</td>
<td>-.01</td>
<td>.03</td>
</tr>
<tr>
<td><strong>Self-interest by (\psi_3)</strong></td>
<td>-.08**</td>
<td>-.15**</td>
<td>-.02</td>
<td>-.15**</td>
<td>-.06**</td>
</tr>
<tr>
<td><strong>Self-interest by (\psi_4)</strong></td>
<td>-.05*</td>
<td>-.07*</td>
<td>-.03</td>
<td>-.04</td>
<td>-.01</td>
</tr>
<tr>
<td><strong>Self-interest by (\psi_5)</strong></td>
<td>.07*</td>
<td>.08*</td>
<td>.04</td>
<td>.03</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>.02</td>
<td>-.02</td>
<td>.03</td>
<td>.04</td>
<td>.02</td>
</tr>
<tr>
<td>(R^2)</td>
<td>.49</td>
<td>.08</td>
<td>.64</td>
<td>.36</td>
<td>.77</td>
</tr>
</tbody>
</table>

| **Other-interest**      |       |            |                  |        |             |
|                         |       |            |                  |        |             |
| \(\psi_1\): honest vs fake good | - .62** | -.10** | -.73** | .57** | - .69** | - .62** | - .74** |
| \(\psi_2\): honest vs fake bad  | .71** | .11** | .84** | -.50** | .99** | .96** | .95** |
| \(\psi_3\): self vs informant  | -.03 | -.06 | -.01 | .07* | -.05** | -.04** | -.05** |
| \(\psi_4\): good x self/informant (\(\psi_1 \times \psi_3\)) | .00 | -.09* | .04 | -.17** | -.03 | -.05* | .01 |
| \(\psi_5\): bad x self/informant (\(\psi_2 \times \psi_3\)) | .05 | .17** | -.02 | .10** | -.02 | -.03 | .00 |
| **Other-interest by \(\psi_1\)** | -.06 | .00 | -.07** | -.06 | -.05** | -.06** | -.03 |
| **Other-interest by \(\psi_2\)** | .01 | -.09* | .06** | -.06 | .06** | .05** | .06** |
| **Other-interest by \(\psi_3\)** | .06** | .06 | .05* | .01 | .03 | .02 | .04*
| **Other-interest by \(\psi_4\)** | .04 | .03 | .03 | .03 | .03 | .04 | .01 |
| **Total** | .04 | .06 | .02 | .05 | -.00 | -.00 | -.01 |
| \(R^2\) | .47 | .06 | .63 | .34 | .77 | .72 | .75 |

* \(p < .05\), ** \(p < .01\)
**Other-interest as a moderator of faking.** As shown in Table 3.4, after controlling for instructions (honest, fake good and fake bad), target (self or other), and the instruction/target interactions, other-interest was not significantly related to any Dark Triad domain or facet when rating a friend (informant-report).

For grandiose narcissism, the relationship between other-interest and dark triad scores was different for the “answer honestly” and “fake bad” conditions. Specifically, there was a stronger positive association of other-interest with narcissism in the “answer honestly” condition compared to the “fake good” condition.

In contrast, other-interest significantly moderated the extent to which people “fake bad” compared to “answer honestly” such that other-interest had a stronger effect for the “fake bad” condition compared to the “honest” condition for vulnerable narcissism, and total psychopathy (but not the other traits). Other-interest also significantly moderated the extent to which people “fake good” compared to “answer honestly” on vulnerable narcissism, total psychopathy, and primary psychopathy suggesting other-interest has a stronger effect on dark triad scores for the “honest” condition than the “fake good” condition.

The beta-weights testing whether other-interest affected whether faking was more pronounced on informant-ratings than self-ratings were not significant for any trait. Thus, there was no support for Hypothesis 4b.

**3.3.3 Discussion**

The results of study 1 demonstrated three things. First, people can fake on self-report measures of the Dark Triad when instructed to do so. The magnitude of faking on the Dark Triad is substantial in almost all cases (all but grandiose narcissism) and is particularly pronounced for faking bad on the psychopathy domain and facets. Second, informants fake at least as much as people rating themselves (and possibly more, in the case of faking good on
psychopathy). Third, on some (but not all) Dark Triad domains and facets, people with high self-interest scores also have higher Dark Triad scores.

As expected, test-takers were able to intentionally fake on self-report measures of the Dark Triad. People “faked good” with very large effects for all Dark Triad domains and facets except for grandiose narcissism. People “faked bad” with moderate to very large effects for all Dark Triad domains and facets. Faking on psychopathy and its facets was consistent with prior meta-analytic results (Walker et al., 2020). The grandiose narcissism results were also consistent with Young (2018) such that when instructed to fake good, people’s grandiose narcissism scores increased; though not significant, this was in the opposite direction to our hypothesis. The null findings for grandiose narcissism may be related to the fact that the items appear socially desirable meaning even those answering honestly may have inflated their scores without realising it was measuring grandiosity. As a result, the ‘honest’ scores obtained on the grandiose narcissism facet were not significantly different to the ‘fake good’ scores on the grandiose narcissism facet. Findings were also in an unexpected direction for Machiavellianism, where instructions to fake-good resulted in higher scores, and instructions to fake-bad resulted in lower scores. This was consistent across both self- and informant-reports, and is inconsistent with prior research on ‘faking good’ on Machiavellianism (Walker et al., 2020).

The items measuring both grandiose narcissism and Machiavellianism may have impacted the direction and degree to which people faked. Some grandiose narcissism items on the Pathological Narcissism Scale (PNI; Schoenleber et al., 2015) might appear to be socially desirable characteristics for someone to possess. For example, “Sacrificing for others makes me the better person” or “I often fantasize about being rewarded for my efforts” both seem to involve an element of positive behaviour (i.e., sacrificing oneself for another, making rewardable effort) as well as the element of recognition/glory as the motivator for such
behaviour (the presumed narcissism-relevant content). Arguably, the positive behaviours may themselves may represent desirable characteristics. The pattern of responding in the current study suggests that when instructed to fake good, people tended to give themselves higher grandiose narcissism scores rather than lower scores. This may be explained by the positive sounding nature of the scale items.

Similarly, the Five-Factor Machiavellianism Scale (FFMI; Collison et al., 2018) is composed of items based on Big Five personality traits and characteristics theoretically associated with the aversive aspects of Machiavellianism. That is, items such as “I like to carefully consider the consequences before I make a decision” or “I like to map out projects before I begin” are meant to represent the long-term strategizing characteristic of high Machiavellian individuals (Collison et al., 2018). In contrast, items such as “I am not an ambitious person” may sound like a negative item for the test-taker to endorse even though this item is meant to assess lower levels of Machiavellianism. Consequently, when instructed to fake good, test-takers’ Machiavellianism scores increased. Conversely, when instructed to fake bad, test-takers’ scores decreased. Considering the Five-factor Machiavellianism Inventory has situated Machiavellianism within the Five-factor model, it is unsurprising the pattern of results found in this study are congruent with patterns of faking found in the Big Five (Viswesvaran & Ones, 1999). That is, we know people give themselves higher scores on Big Five traits when faking good than when answering honestly. It is also possible that the positively valanced nature of the grandiose narcissism and Machiavellianism items could explain the unexpected direction of these results.

The concept that test-takers are influenced by how positive or negative an item sounds is not a new observation. In fact, the frequency at which items are endorsed can be predicted from the socially desirable nature of the items themselves (Edwards, 1953, 1957; Norman, 1967). As is typical of personality scales in general, the Dark Triad items are endorsed on a
ratings scale in which one end of the pole is usually more desirable than the other (Saucier, 1994). The very large effects of informant-report faking in the present study may be influenced by the socially desirable, evaluative nature of the items.

### 3.4 Study 2: Expert Ratings of Evaluative Content in Dark Triad Measures

Study 2 was designed as a small follow up study to clarify and test whether the results of Study 1 can be explained by the social desirability of the items which make up each scale. That is, Study 2 aimed to determine if the evaluative content of the items from Study 1 measures underpinned the direction and extent of faking. The broad aim of this study is to generate item-level and scale-level indices of evaluative content (or social desirability) that can be used to empirically test the extent to which faking on each scale (and each item) is due to the evaluative content of the items. Anglim et al. (2017) and Bäckström et al. (2009) found that people decide to endorse items based on how positive or negative the item sounds. As such, we expect there will be significant associations between the extent of faking in Study 1 and the social desirability estimates produced in study 2.

**Hypothesis 1:** The direction and extent of faking on each scale in Study 1 will be **positively associated with** the social desirability ratings. Specifically, among the seven scales/subscales (three domains and four facets), the effect sizes comparing “honest” and “fake” (Cohen’s $d$) will be significant and highly correlated with the median social desirability rating of expert raters. These values are expected to be positive for “fake good” (Hypothesis 1a) and negative for “fake bad” (Hypothesis 1b).

**Hypothesis 2:** The direction and extent of faking will be strongly associated with the social desirability of each item. For each subscale, the degree of faking on each item will be strongly predicted by the item's social desirability (as rated by experts). We expect this hypothesis to hold for both faking good (Hypothesis 2a) and faking bad (Hypothesis 2b) from study 1.
3.4.1 Method

Participants

We recruited nine experts working in related research fields (e.g., personality and forensic psychology research). Of the nine experts, 33% had been granted a PhD and 66% were granted an undergraduate psychology degree and currently completing a PhD. The experts were 77% female.

Materials and Procedure

Participants were presented with the same Dark Triad measures in Study 1. Instead of completing the measures to rate their own personality characteristics, participants were instructed to rate each item from 1-7 with 1 = “extremely socially desirable” 4 = “neither socially desirable or undesirable” and 7 = “extremely socially undesirable”.

3.4.2 Results

Table 4 shows the reliability and descriptive statistics of the expert social desirability ratings of each scale. Inter-rater reliability was excellent for Machiavellianism, total psychopathy and the facets of psychopathy (ICC = .94 to .97) and was somewhat lower but still good for narcissism and its facets (ICC = .77 to .80). Across all items, the mean social desirability was the most extreme for Machiavellianism ($M = 4.10$, the mid-point of the scale was 4), followed by vulnerable, total, grandiose and narcissism ($M = 4.94$, 4.71, and 4.42 respectively), and finally primary, total, and secondary psychopathy ($M = 6.00$, 5.70, and 5.20 respectively). The scales social desirability mean ratings correspond to the pattern of faking seen on the different scales in study 1. That is, the most desirable scale (Machiavellianism) shows a pattern of increasing under fake good (and decreasing under fake-bad). The least desirable scales (psychopathy and its facets) show the greatest degree of faking.
Hypothesis 1: Scale desirability ratings. For the 7 scales/subscales of the dark triad \((n = 7)\), we calculated the correlation of \(d\)-scores with expert median ratings of social desirability. Results demonstrate that scale-level social desirability ratings are negatively correlated with the ‘answer honestly’ – ‘fake good’ score differences for both self-reports \((r = -.92, p = .002)\), and informant-reports \((r = -.91, p = .002)\). The more socially desirable scales showed greater score decreases under ‘fake good’ instructions. The scale level social desirability rating was positively correlated with the ‘answer honestly’ – ‘fake bad score differences for both self-reports \((r = .69, p = .043)\), and informant-reports \((r = .79, p = .017)\). The more socially desirable scales showed greater score increases under ‘fake bad’ instructions.

Hypothesis 2: Item desirability and “faking good”. For each of the seven scales, we tested whether there was a significant association between item-level social desirability
(median expert social desirability rating of each item) and the extent of faking in Study 1 (i.e., the difference in item means for the ‘fake good’ minus ‘answer honestly’ and ‘fake bad’ minus ‘answer honestly’ conditions). The results of this analysis are presented in Table 3.5 which shows the correlations of the item social desirability with the extent to which items were faked good or faked bad for self-ratings and informant ratings. Each item was given a social desirability estimate based on the median social desirability expert rating across the 9 experts. The median was chosen to eliminate outlier influence. Pearson correlations were calculated for self-report honest versus faking conditions for each of the Dark Triad domains and facets. For example, for the \( n = 52 \) Machiavellianism items, the correlation between the item social desirability estimates and ‘fake good’ versus ‘answer honestly’ item differences for self-ratings was \( r = -.52 \), indicating that stronger evaluative content leads to greater ‘faking good’ on Machiavellianism items.

For faking bad, there were very strong positive correlations between social desirability estimates and the extent of faking bad for all dark triad traits. These results were consistent across both self-reports (\( r = .77 \) to .98) and informant-reports (\( r = .65 \) to .98). Results were clearest for psychopathy and its facets, with correlations above .90 for both self-reports (\( r = .90 \) to .98) and informant-reports (\( r = .94 \) to .98).

For faking good, the correlations between social desirability item estimates were consistently significant and negative for Machiavellianism and psychopathy (\( r = -.52 \) to -.78 for self-ratings and -.62 to -.94 for informant-ratings). However, associations were not significant for narcissism (either the total score or its facets) for self-ratings (\( r = -.15 \) to .41), and were not significant for the two facets of narcissism for informant ratings (\( r = -.26 \) and -.48). We tested whether the associations of social desirability/faking correlations were significantly stronger for faking bad versus faking good (considering only the magnitude of correlations, as effects were in opposite directions). For self-reports, social desirability
estimates were a significantly stronger predictor of faking bad than faking good for all scales except for secondary psychopathy. For informant-ratings, social desirability estimates were a significantly stronger predictor of faking bad than faking good only for total psychopathy (and for no other domain or facet). Taken together, these results indicate that item-level social desirability is a key determinant of the extent of response distortion, particularly for faking bad.

Table 3.5  
Correlations of expert social desirability ratings and participant's mean change in faking for self-reports, and informant-reports.

<table>
<thead>
<tr>
<th></th>
<th>Self-Report</th>
<th>Informant-Report</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fake good</td>
<td>Fake bad</td>
</tr>
<tr>
<td></td>
<td>- Honest</td>
<td>- Honest</td>
</tr>
<tr>
<td>Total Narcissism</td>
<td>-.15</td>
<td>.87**</td>
</tr>
<tr>
<td>Grandiose</td>
<td>.05</td>
<td>.90**</td>
</tr>
<tr>
<td>Vulnerable</td>
<td>.41</td>
<td>.88**</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Machiavellianism</td>
<td>-.52**</td>
<td>.77**</td>
</tr>
<tr>
<td>Total Psychopathy</td>
<td>-.57**</td>
<td>.95**</td>
</tr>
<tr>
<td>Primary</td>
<td>-.84**</td>
<td>.98**</td>
</tr>
<tr>
<td>Secondary</td>
<td>-.78**</td>
<td>.90**</td>
</tr>
</tbody>
</table>

Note. z = calculation of z-score to assess whether the magnitude of correlations (ignoring direction) for ‘fake good’ versus ‘fake bad’ is significantly different.

* p < .05, ** p < .01
3.4.3 Discussion

Results from Study 2 clearly demonstrate that the social desirability of items is not equal across the three Dark Triad scales, and that the extent of faking is highly dependent on the social desirability of the items (especially for ‘faking bad’). As expected, there was high agreement between experts on the evaluative content of the items. According to the expert ratings, a higher proportion of items were rated as neutral compared to Machiavellianism and psychopathy, which more clearly demarcate socially desirable/undesirable items. Overall, the Dark Triad scale items contain highly socially desirable/undesirable items. As such, it is unsurprising the evaluative nature of the items can influence how much, or how little, an individual is willing to endorse those items.

The results of Study 2 demonstrate that stronger evaluative content leads to greater faking when measuring Dark Triad domains and facets. When faking good, it appears test-takers use the positively valanced evaluative content to produce larger effect sizes for both self- and informant-reporting. In contrast, when faking bad, test-takers appear to use the negatively valanced evaluative content to produce larger effect sizes for both self- and informant-reports. While an intuitive finding, it also provides empirical evidence suggesting test-takers can identify and use the evaluative content of Dark Triad scales to choose how to distort their responses appropriately based on a specific goal (in this case, faking good or faking bad). However, Machiavellianism was an exception.

The expert ratings of social desirability may help to explain the unexpected results in Study 1 for Machiavellianism and grandiose narcissism. However, unlike the other Dark Triad domains and facets, faking on Machiavellianism was in the opposite direction to that predicted in Study 1. That is, when instructed to fake good, people tended to increase their Machiavellianism scores, compared to faking bad, where people decreased their scores. The evaluative nature of the item content may provide some insight. The results of Study 2 show
that expert ratings of the FFMI consists of equal parts positively and negatively valanced items, with approximately 21% of the items rated as neutral. Compared to the other Dark Triad scales, there is a high proportion of socially desirable items. Accordingly, when instructed to fake good, the test-taker may have perceived the items as measuring something positive and decided to endorse those items accordingly. This is consistent with the findings of Study 2 which found Machiavellianism to have the highest social desirability rating among the Dark Triad domains and facets.

As expected, there were large correlations between the extent to which people faked and the social desirability of the Dark Triad rating scale measures and items. This demonstrates that people are using the items' evaluative content rather than considering the match between their own personality and descriptive content of the items when in the faking conditions. This is consistent between both self- and informant-report ratings. Additionally, these results show that test-takers use the evaluative content significantly more when they fake-bad than when they fake-good. Importantly, the power to detect a significant effect will be different for different scales due to the different numbers of observations and items per scale. However, the effect sizes in this case, rather than the significance, are the critical feature. Insofar as faking good is concerned, the results are concordant with Anglim et al., (2017) and Bäckström et al.’s (2009) findings that people choose to endorse items based on how positive or negative the item sounds.

3.5 General Discussion

Concern relating to the susceptibility of self-report scales to response distortion is not a new observation. Since 1957 when Edwards suggested personality scales may be measuring social desirability rather than the personality construct itself, the field of personality research has investigated response distortion on self-report personality scales. While this research has primarily focused on the Big Five/FFM models of personality (MacCann, 2013), the present
studies sought to examine this within the context of the Dark Triad of personality. As the use of self-report rating scales have been criticized for potential response distortion, the use of informant-reports became a popular way to confirm the accuracy of self-reported personality traits (Kim et al., 2019; Vazire, 2006). A recent meta-analysis examining self-other agreement concluded the high convergence between self- and informant-reports was evidence for limited response distortion on self-report scales (Kim et al., 2019).

Kim et al.’s (2019) conclusions appear to confirm a widely held belief that informants do not engage in socially desirable responding. Based on the high-level of self-informant agreement, Kim et al., did not appear to consider that the informant may have: (a) endorsed their responses based on the evaluative content of the items, or (b) that the degree of friendship biases in informant-report ratings may be similar to self-report response distortion. Although, based on the current findings which show self-report and informant-reported Dark Triad scores are heavily influenced by the evaluative content, it is possible the pattern of response distortion is consistent between self- and informant-report ratings. That is, Dark Triad domain and facet scores are increased on positively valanced items and decreased on negatively valanced items. In other words, high self-other agreement may not represent agreement on a personality trait per say, but instead, agreement on the social desirability of the items they are endorsing. Substantial research examining the extent to which the evaluative and descriptive content of both self- and informant-report measures influence socially desirable responding and how this impacts self-informant report measures is required.

Although Study 1 was a lab-based instructed faking study, the findings from both Study 1 and 2 have relevance beyond a research context. However, it should also be noted that a limitation of the current research is the presentation of the fake bad manipulation. Test-takers were instructed to answer as if they (or their friend) did not want the job. This element
of the design limits the generalizability of this study to other fake bad/malingering studies. It is likely presenting an exaggerated ‘bad’ profile would be used in situations where the exaggeration would produce a more desirable outcome such as a more lenient prison sentence (Rogers et al., 2003) as compared to wanting to avoid receiving a job offer.

Assessment of Dark Triad traits has gained popularity following the highly publicized scrutiny surrounding organizational misconduct (Van Scotter & Dea Roglio, 2018), police misconduct (Semrad & Scott-Parker, 2020), the abuse and murder of prisoners by armed forces and CIA operatives (Bartone, 2010) right through to the Dark Triad predicting workplace outcomes including job performance, citizenship behavior, counterproductive workplace behavior, and leadership styles (Spain et al., 2014). As assessment of Dark Triad traits becomes more prevalent in organizations interested in screening out individuals possessing maladaptive traits during the selection process (Spain et al., 2014), confidence in these measures to provide accurate and reliable assessment is paramount. This is particularly true in situations where the applicants are responsible for or have access to, vulnerable communities.

This study provides the first comprehensive assessment of instructed faking investigating the extent to which faking can occur on both self- and informant-reports of the Dark Triad. The current data confirm that people can substantially fake on self-report measures of the dark triad. One of the more significant findings to emerge from this study is that informants can also fake on behalf of their target. Before this study, evidence that informants are unlikely to fake was purely anecdotal. Additionally, these results have shown that the dark triad measures contain a substantial amount of evaluative content. Taken together, these results demonstrate the importance of closely examining the extent to which response distortion is relevant for self- and informant-reports, as well as the impact of evaluative content on item endorsement. If people self-report themselves favorably and
informants also rate their targets favorably, then the Dark Triad assessment utility is diminished. Although self-report scales continue to undergo intense scrutiny regarding their accuracy as a result of response distortion there has not been the same scrutiny for informant-reports. This study has established that informants can fake and lays the foundation for future research to continue to explore informant faking. In particular do informants fake in a high-stakes real-world context, and what are the implications for score interpretation if they do fake. In order for the confidence in informant-reports to continue, a similar level of scrutiny must be applied to examining the psychometric properties of informant-reports on the Dark Triad of personality.
References


Karpman, B. (1941). On the need of separating psychopathy into two distinct clinical types: the symptomatic and the idiopathic. *Journal of Criminal Psychopathology, 3*, 112-137


CHAPTER 4: General Discussion

Self-report rating scales are commonly used to assess people’s personality profiles but have been found to be susceptible to faking (Viswesvaran & Ones, 1999). Informant-reports have been identified as another method of obtaining information about a person’s personality and have often been used as a means by which to confirm the accuracy of self-reported personality scores (Kim et al., 2018). While there is evidence that people can and do fake on self-reported personality measures (Birkeland et al., 2006; Viswesvaran & Ones, 1999), it has almost always been assumed that informant-reports are not susceptible to faking. The broad aim of this thesis was to: (a) investigate the extent to which people distorted their responses on self-report measures of the Dark Triad, (b) to investigate whether informants can also distort their responses, (c) the extent to which self-interest or other-interest may influence how much people fake for themselves, or others, and (d) the extent to which the evaluative content of the Dark Triad items has an effect on item endorsement. A summary of Chapters 2 and 3 will be provided in the paragraphs below.

In the meta-analysis and systematic review reported in Chapter 2, we examined the extant studies that analyzed faking good and faking bad on Dark Triad measures. The meta-analysis and systematic review demonstrated that people can significantly fake good and fake bad on psychopathy and Machiavellianism measures. That is, when instructed to ‘fake good’, people can lower their psychopathy scores, and when instructed to ‘fake bad’, people can increase their psychopathy scores. The single narcissism study included in the systematic review revealed that people can also fake on narcissism, though faking good on narcissism was in the opposite direction than expected. That is, narcissism scores increased when faking good instead of decreasing as expected. Although the magnitude of faking effects were consistent with prior Big Five meta-analytic findings (Viswesvaran & Ones, 1999), this meta-analysis was limited by the small number of studies available to analyze. This was particularly pronounced for Machiavellianism and narcissism which were unable to be meta-
analyzed due to the small sample of available studies. Nevertheless, the significant effects of faking good and faking bad on psychopathy measures alongside evidence of faking for narcissism and Machiavellianism provided an initial impetus for hypothesising that self-report Dark Triad rating scales are susceptible to faking.

Chapter 3 described two studies. The first study (the experimental design) directly tested the extent to which people fake on self- and informant-report measures of the Dark Triad. This study found that: (a) that when instructed to, people can fake on self-report measures of the Dark Triad which is consistent with the meta-analytic and systematic review findings in Chapter 1, (b) the extent to which people faked was considerable on almost all Dark Triad domains and facets (except grandiose narcissism). This was particularly evident for faking bad on the psychopathy domain and facets, (c) the extent to which informants fake is at least as marked as people faking for themselves which suggests that contrary to prior assumptions, informants can fake on personality scales – in this case, Dark Triad scales, and (d) that the Dark Triad domains and facets (except for secondary psychopathy) were positively associated with self-interested motivation.

While the direction of faking in Study 1 was consistent with expectations across almost all of the Dark Triad domains and facets, the direction of faking for Machiavellianism was contrary to expectations. In contrast to the systematic review findings from Chapter 2, results from Study 1 (Chapter 3) demonstrate that when instructed to fake good on Machiavellianism, test-takers increased (rather than decreased) their scores, and when instructed to fake bad on Machiavellianism, test-takers decreased (rather than increased) their scores. A possible explanation for this might stem from the development of the Five-factor Machiavellianism Inventory (FFMI) as situated within the FFM framework. For example, the FFMI shows large positive correlations with several facets of Conscientiousness, with the largest for Self-Discipline, Competence, and Achievement-Striving ($r = .44$ to $ .48$; Kückelhaus et al., 2020). Of the six conscientiousness facets, these three (i.e., those with the
strongest associations with the FFMI) also show the largest score increases under instructions to fake good (with effect sizes of 1.16 to 1.58; Ziegler et al., 2010). As such, it is not surprising that the FFMI shows a pattern of faking incongruent with other Dark Triad scales but consistent with faking on Big Five measures. That is, scores increase when people fake good, and decrease when people fake bad (Viswesvaran & Ones, 1999; note that neuroticism reversed so as to be ‘emotional stability’). Where Dark Triad researchers are usually concerned about test-takers suppressing their Dark Triad scores, these findings, while preliminary, suggest that scores on the FFMI may be inflated rather than suppressed when there is motivation to distort responses.

The second study (Study 2) in Chapter 3 tested whether the evaluative content of the Dark Triad measures influenced the extent to which people faked for themselves and others. This was achieved by using expert raters who were tasked with rating the social desirability/undesirability of each item included in the Dark Triad scales used in Study 1 (Chapter 3). There were clear differences in the social desirability of item content across the scales. The FFMI items were relatively neutral on average (with expert ratings close to the mid-point for social desirability) whereas the primary psychopathy items were relatively undesirable on average (with expert ratings close to the ‘undesirable’ pole, at 6.00 on a 7-point scale). These differences mapped onto the extent of faking, both at the scale level (i.e., scores changes to scales under instructed faking) and at the item level (i.e., scores changes to items under instructed faking). However, the social desirability of items was more strongly linked to faking bad than on faking good, particularly in the case of self-reports. It is not surprising that the extent to which people fake on Dark Triad scales is highly dependent on the social desirability of the items (especially for ‘faking good’). The results of Study 2 demonstrated that faking was more marked when the scale items contained a high degree of evaluative content.
The results from Study 2 may also provide some explanation for the unexpected direction of faking effects found for Machiavellianism in Study 1. Compared to the other Dark Triad scales, the FFMI has a high proportion of positively valanced items as rated by the expert raters. Consequently, when instructed to fake good, the test-taker may have endorsed the items based on their perception that the items were measuring desirable qualities. Overall, the results from Study 2 suggest test-takers can identify, and use, the evaluative content of the Dark Triad scales to decide how to distort their responses according to a specific goal (i.e., faking good or faking bad).

4.1 Implications for Research and Theory

Implications for Self-Report Measures of the Dark Triad. The results of this thesis provide a clear indication that, under faking conditions, people can fake on measures of the Dark Triad. Furthermore, the results suggest that the evaluative content of Dark Triad scales strongly influences the extent to which people distort their responses. One of the issues that emerges from the findings of this thesis relates to the application of Dark Triad assessment. Where the Big Five personality traits are often assessed as part of a job selection process, Dark Triad assessments are often used in forensic settings where high-stakes outcomes for the test-taker may be particularly consequential (e.g., psychopathy is measured as part of presentencing investigations; Edens, 2001). Given the high-stakes contexts of such situations, it is reasonable to assume people are likely to fake. Therefore, the extent to which faking can alter Dark Triad scores is a critical consideration. This is especially necessary given the increasing popularization the Dark Triad, and the subsequent use of Dark Triad scales as assessment tools to ‘screen out’ people with psychologically deviant tendencies during the recruitment process for law enforcement and armed services. As the evidence currently stands, caution should be taken when interpreting self-reported dark triad assessment scores as the basis for high-stakes decision making.
Implications for Informant-Report Measures. It has been suggested informant-report rating scales are not susceptible to response distortion unlike self-report scales (Kim et al., 2018). This does not appear to be the case. The findings of this thesis, while preliminary, suggests that the use of informant-reports may not be an ideal method to determine the accuracy of self-report rating scales. Both self- and informant-reported Dark Triad scores were found to be strongly influenced by the evaluative content of the Dark Triad scales. As such, it is possible the pattern of response distortion is consistent between self-report and informant-report ratings such that the perception of socially desirable items is similar across self- and informant-reports. That is, the agreement found in self-other ratings may represent agreement on the social desirability of the items as opposed to agreement on a particular personality trait or set of traits. These preliminary findings demonstrate that informants can fake on the Dark Triad, but they do not indicate whether informants do fake on the Dark Triad in applied contexts. Although further research is needed to determine whether informants do fake, it is noteworthy that prior Big Five research has found people do fake their self-reported personality scores (Birkeland et al., 2006).

Implications for the Evaluative Content of Dark Triad Scales. These findings broadly support the work of prior studies which have found people tend to endorse items based on their evaluative content (Anglim et al., 2017; Bäckström et al., 2009). In this case, the Dark Triad scales are particularly socially undesirable. According to the findings of this thesis, we can infer that people may suppress their narcissism and psychopathy scores and inflate their Machiavellianism scores based on the items' social desirability, thereby limiting the ability to interpret Dark Triad scores accurately. As a result, the utility of these scales in an applied context is diminished at best, or untrustworthy at worst.

4.2 Limitations and Future Directions

An important question that remains unanswered by the current research is the extent to which informants do fake. That is, the instructed faking studies examine faking in an
artificial or simulated context where people are encouraged to fake. They therefore more strongly represent a statement about the extent to which faking is possible (rather than the extent to which faking likely occurs in practice). There is meta-analytic evidence that individuals do fake on self-report Big Five scales in high-stakes settings (Birkeland et al., 2006), but there is no complementary literature for informant-reports. While this thesis provides preliminary evidence that informants can fake on Dark Triad measures, further research should be undertaken to investigate whether informants do fake in practice, and the extent to which they fake. Future studies should be expanded beyond the Dark Triad traits and applied to personality more broadly. Essentially, the rigour applied to examining the extent to which people fake on self-reports, and the implications of that faking should be applied to informant-report faking. This is essential in order to confidently use informant-reports as an additional source of information about another person in applied high-stakes settings. The idea that informants are not affected by the stakes of the assessment (because the outcome applies to the target not the informant) may not always be true for several reasons. First, applicants will usually be asked to nominate informants. They would generally select who they trust like them and want them to succeed. As such, the “stakes” for the informant are whether a person they like (a colleague, direct report, student or even personal friend) receives a valued outcome. Second, the cost of faking (e.g., the awkwardness of being questioned about the potential truth of a rating or questioned about a lack of modesty) may also be lower for informant-reports than self-reports. The informant never has to deal directly with the organization they have deceived, whereas the self will (if all goes well and they get the job). Third, there are diffuse positive outcomes for the individual or their organization if their supervisees or students succeed. For example, the prestige of a private preparatory school may partly derive from the proportion of students who get into high-ranking careers like law or medicine. A teacher or guidance counsellor would be motivated, in such a case, to fake good on a rating scale so that their students obtain entrance to a valued course or career.
Several questions remain unanswered as to the unexpected direction of faking produced using the Five-factor Machiavellianism Inventory. While the evaluative content was shown to have influenced the extent to which test-takers endorsed positively and negatively valanced items, it is unclear whether the degree to which faking occurred in the opposite direction to the other Dark Triad traits was due to the scale being situated within the FFM framework or not. That the evaluative content of the Dark Triad influences the degree to which people endorse items is an important issue for future research. Considering the Dark Triad scales necessarily sound ‘dark’, it is reasonable to assume that on average, people rate themselves (and others) more positively. As has been achieved with Big Five scales (Bäckström et al., 2009), future research might consider the opportunity to develop Dark Triad scales containing less evaluative content.

Further, examining the transparency of the items including in Dark Triad rating scales is needed. In Paulhus and Vazire (2007) a distinction was made at between personality items where measurement is direct (i.e., face validity; e.g., “I often fantasize about being rewarded for my efforts”) versus indirect (e.g., “If I ruled the world, it would be a better place” is not in fact a measure whether the world would be a better place if the respondent ruled it”). Paulhus and Vazire suggested that items hiding/masking the true intention to measure a specific construct were indirect measures, and items that specifically asked for the test-taker’s opinion about the construct directly were direct measures. Additionally, it would be beneficial to examine the multiple different measures of the dark triad, to disentangle construct effects from test effects. This examination is important in order to account for how much variance in faking is accounted for by the characteristics of the measure (specifically, the social desirability of the items/scale).

This distinction between transparent and non-transparent items in personality assessment may moderate both the extent of faking and the extent to which faking differs across self- and other-reports. Specifically, this factor (item transparency or the direct/indirect
measurement distinction) may be an additional way in which faking can be reduced. That is, if the construct itself is socially undesirable (which arguably is true for the dark triad traits), it may not be possible to adequate to develop “neutral” items as per Bäckström et al.’s (2009). Instead, it may be necessary to conceal the true nature of the construct from the test taker.

Prior research has examined implicit testing methods of measuring personality, including ‘dark’ traits such as aggression (e.g., James & LeBreton, 2010; Friedman et al., 2020). However, like faking research, little to none of this research has examined informant-reports. A fruitful future direction for research would be to examine this type of implicit measurement as a way of assessing dark triad traits, including the extent to which they are fakeable under both self- and informant-reports.

Finally, the sample used in Study 1 (Chapter 3) consisted of undergraduate psychology students examining only one faking context (job applicant). To develop a full picture of the extent to which people fake for themselves, and others on the Dark Triad (and personality more broadly), additional studies will be needed that: (a) replicate these results in a broader community sample group, (b) systematically examine real-world influences (e.g., job type, job desirability, job market, education admissions, competitiveness of degree programs etc.), and (c) consider the underlying motivation of what drives faking on self- and informant-reports. Further from the final point (c), future research should consider how informants choose their targets, while actively working toward eliminating any anchoring biases that may be present. For example, it is unclear in these results whether the informant was truly responding about their target, or whether the informant’s focus had shifted toward another anchor (such as themselves, or another person).

4.3 Concluding Statements

Self- and informant-reports are the most commonly used methods to obtain information about someone’s personality. That people fake on self-report measures is understood (Birkeland et al., 2006; Viswesvaran & Ones, 1999). One of the more important
findings to emerge from this thesis is that informants can also fake on behalf of someone else. This thesis has extended investigations of faking on self-report measures of the Dark Triad by examining the evaluative content of the Dark Triad scales, and provides the first comprehensive assessment of instructed faking investigating the extent to which people can also fake for someone else. If both self- and informant-reports are rated favourably, then the assessment utility of Dark Triad scales are diminished. Results from this thesis demonstrates that informants can fake laying the groundwork for future work to explore informant faking. In particular, whether informants do fake in a high-stakes real-world context, and the implications for score interpretation if they do. In order to maintain confidence in the utility of informant-reports, a similar level of scrutiny must be applied to examining the psychometric properties of informant-reports not only on the Dark Triad, but personality more broadly.
REFERENCES


Karpman, B. (1941). On the need of separating psychopathy into two distinct clinical types: the symptomatic and the idiopathic. *Journal of Criminal Psychopathology, 3*, 112-137


APPENDIX A

Instructions adapted from Heggestad et al. (2006)


NOTE: As this is a between-person design, each participant was randomly allocated to ONE of the following SIX instructional conditions (with randomization programmed into Qualtrics)

Condition 1: Instructions for “Honest about SELF”

As you complete the following questionnaires,

On the next screens, you will complete a personality rating scale and a situational test of personality. For both of these tests, respond by describing yourself as accurately and honestly as possible.

Condition 2: Instructions for “Honest about OTHER”

Think of a friend you have (who is not your partner) who is about the same age and gender as yourself.

What was the gender of your friend? MALE FEMALE OTHER

What is the age of your friend? ______ years

On the next screens, you will complete a rating scale and a situational test that ask about your friend’s personality. For both of these tests, respond by describing your friend as accurately and honestly as possible.

Condition 3: Instructions for “Want the job for SELF”

On the next screens, you will complete a personality rating scale and a situational test of personality.

As you complete these tests, we want you to pretend that you are a job applicant for a job that you really want. Take a minute to think about a job that you think would be great. Imagine that the interviewer tells you that before she can hire you, you need to complete a couple of questionnaires to see if you really are good fit for the company. If you score well, you will get the job. If you do not score well, you will not get the job. So, as you complete the next two questionnaires, we want you to respond to each statement by describing yourself with the goal of making the company believe that you will make a good employee.
Condition 4: Instructions for “OTHER wants the job”

Think of a friend you have (who is not your partner) who is about the same age and gender as yourself.

What was the gender of your friend? MALE FEMALE OTHER

What is the age of your friend? ______ years

On the next screens, you will complete a rating scale and a situational test that ask about your friend’s personality.

As you complete these tests, we want you to pretend that your friend is a job applicant for a job that they really want. Take a minute to think about a job that your friend really wants. Imagine that the interviewer tells your that before the company can hire them, they need to ask a referee to complete a couple of questionnaires to see if your friend will be a good fit for the company. If your friend scores well, your friend will get the job. If your friend does not score well, your friend will not get the job. So, as you complete the next two questionnaires, we want you to respond to each statement by describing your friend with the goal of making the company believe that your friend will make a good employee.

Condition 5: Instructions for “Do NOT want the job for SELF”

On the next screens, you will complete a personality rating scale and a situational test of personality.

As you complete these tasks, we want you to pretend that you are a job applicant for a job that you really do NOT want. Take a minute to think about a job that you think would be terrible. Imagine that the interviewer tells you that before the company can hire you, you need to complete a couple of questionnaires to see if you really are good fit for the company. If you score well, you will get the job. If you do not score well, you will not get the job. So, as you complete the next two questionnaires, we want you to respond to each statement by describing yourself with the goal of making the company believe that you will NOT make a good employee.

Condition 6: Instructions for “OTHER does NOT want the job”

Think of a friend you have (who is not your partner) who is about the same age and gender as yourself.
What was the gender of your friend?  MALE   FEMALE   OTHER

What is the age of your friend? ______ years

On the next screens, you will complete a rating scale and a situational test that ask about your friend’s personality. As you complete these tests, we want you to pretend that your friend is a job applicant for a job that they really do NOT want. Take a minute to think about a job that your friend would think was terrible. Imagine that the interviewer tells you that before the company can hire them, they need to ask a referee to complete a couple of questionnaires to see if your friend will be a good fit for the company. If your friend scores well, your friend will get the job. If your friend does not score well, your friend will not get the job. So, as you complete the next two questionnaires, we want you to respond to each statement by describing your friend with the goal of making the company believe that your friend will NOT make a good employee.
APPENDIX B:

Manipulation Check (administered after the instruction screen)

The instructions asked me to:
1) complete some intelligence tests as part of a job application
2) complete some intelligence tests as quickly as possible
3) rate my own personality as accurately as I can
4) rate a friend’s personality as accurately as I can
5) complete some personality tests as quickly as possible
6) rate myself as part of a job application for a job I want
7) rate myself as part of a job application for a job I DO NOT want
8) rate my friend as part of a job application for job they want
9) rate my friend as part of a job application for a job they DO NOT want

<If incorrect, then the next screen will say “Please click here to see the instructions again” and then re-route to the instructions>
APPENDIX C

Faking for "Self" vs "Other" on Dark Triad Measures of Personality (#26832)

This pre-registration is not yet public. This anonymized copy (without author names) was created by the author(s) to use during peer-review. A non-anonymized version (containing author names) will become publicly available only if an author makes it public. Until that happens the contents of this pre-registration are confidential.

1) Have any data been collected for this study already?
It's complicated. We have already collected some data but explain in question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?
Our first aim is to test the extent to which instructions to "fake good" and "fake bad" on rating scales of the dark triad of personality will change scores, as compared to instructions to "answer honestly". Our second aim is to test whether the extent of faking differs for self-ratings compared to rating someone else. Our third aim is to test whether personal motivations moderate the extent to which someone will fake for themselves compared to others.

H1: Individuals will fake on all measures of the dark triad. There will be a significant mean difference on all dark triad scores for 'honest' as compared to 'fake good' and 'fake bad' conditions. We expect all dark triad scores to be higher in the 'fake bad' condition and lower in the 'fake good' condition as compared to the 'answer honestly' condition.
H2: Individuals will give better scores to themselves than others. That is, all dark triad scores will be significantly lower for self-reports than other-reports.
H3: There will be an interaction between instruction condition and ratings target. People will show a greater degree of faking for others compared to faking for themselves.
H4: Self/other-interest motivations will moderate the extent to which people fake for themselves compared to others.

3) Describe the key dependent variable(s) specifying how they will be measured.
Scores on dark triad questionnaires.

4) How many and which conditions will participants be assigned to?
This is a 2 x 3 between person design crossing the ratings target (2 levels: self and other) with the instruction condition (3 levels: answer honestly; fake good; fake bad).

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will run a 2-way ANOVA to test the 2 main effects and the interaction effect. For the 3-level variable (instruction condition), we run dummy-coded contrasts with "honest" as the reference group, such that we are comparing "fake good" and "fake bad" to honest". We will run separate ANOVAs for each of the dark triad domains and facets.

To test whether self/other interest values moderate effects, we will run the above ANOVA as a regression (dummy coding the instruction condition with "honest" as the reference group), and add the variable self-other as an additional variable, and also interaction terms with each of the 2 main effects and their interactions. We will run these regressions separately for each of the dark triad domains and facets.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We plan to flag participants' data for exclusion if:
1. They have failed both data check items.
2. They have selected 'not well' or 'not well at all' from the demographic question "how well do you speak English". If significant differences exist at more than chance, participants will be excluded when selecting anything other than 'very well' as this may indicate responses are due to limitations in understanding rather than socially desirable response tendencies.
3. If the time stamp suggests they took more, or less than a third of median response time to complete the study.
4. If their response in any free-text fields do not answer the question or do not make sense. Exclusion decisions will be made on a case-by-case basis depending on number of, and type of flags.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
There will be approximately 1,100 first year psychology students taking part. They are participating as part of an official tutorial exercise forming their assessment project. Students must consent for their data to be used beyond the tutorial activity for research purposes. Only students who provide consent for their data to be used for research purposes will be included in these analyses. Students who do not provide consent will be removed from the data prior to analyses.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
Collection for this study has already commenced. The delay in this pre-registration is due to illness and being unable to upload prior to the start of data collection. There has been no attempt to access the data in order to form alternate hypotheses. Data collection is running from 12/08 - 16/08/19. This document will be uploaded 14/08/19

Verify authenticity: http://aspredicted.org/blind.php?x=ke3wc3
**APPENDIX D: Chapter 3 Data**

The below web address contains the data collected and analysed for Chapter 3, study 1.

https://dev.azure.com/swal6224/_git/MPhil%20Data%20(Sarah%20A.%20Walker)

Data Legend

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SO1 - SO18</td>
<td>Self-Other Interest Inventory items (SOII) (Gerbasi &amp; Prentice, 2013)</td>
</tr>
<tr>
<td>S.H.02 - S.FB.2</td>
<td>Instructions</td>
</tr>
<tr>
<td>FFML.S.CI</td>
<td>Manipulation Check 1</td>
</tr>
<tr>
<td>P.S.C1</td>
<td>Manipulation Check 2</td>
</tr>
<tr>
<td>O.H.2 – O.FB.5</td>
<td>Instructions</td>
</tr>
<tr>
<td>PNI.O.C1</td>
<td>Manipulation Check 3</td>
</tr>
<tr>
<td>P.O.C1</td>
<td>Manipulation Check 4</td>
</tr>
<tr>
<td>DD1 – DD12</td>
<td>Dirty Dozen items 1-12 (Jonason &amp; Webster, 2010)</td>
</tr>
<tr>
<td>FFMI1- FFMI52</td>
<td>Five-Factor Machiavellianism Inventory (FFMI) items 1-52(Collison et al., 2018)</td>
</tr>
<tr>
<td>PNI1 – PNI28</td>
<td>Pathological Narcissism Inventory (PNI) items 1–28 (Schoenleber et al., 2015)</td>
</tr>
<tr>
<td>LSRP1 – LSRP 26</td>
<td>Levenson’s Self-report Psychopathy Scale (LSRP) items 1-26 (Levenson et al., 1995)</td>
</tr>
<tr>
<td>DDMach, DDPsych, DDNarc</td>
<td>Dirty Dozen Total scores for Machiavellianism, psychopathy, narcissism</td>
</tr>
<tr>
<td>FFMIAntag, FFMIAgency, FFMIPlanful, FFMITotal</td>
<td>FFMI domain and facet scores</td>
</tr>
<tr>
<td>PNIGrandiose, PNIVulnerable, NarcTotal</td>
<td>PNI Domain and facet scores</td>
</tr>
<tr>
<td>PNIExploit – PNIER</td>
<td>PNI sub-facets</td>
</tr>
<tr>
<td>LSRPPrimary, LSRPSecondary, PsychTotal</td>
<td>LSRP domain and facet scores.</td>
</tr>
<tr>
<td>Self</td>
<td>(SOII) Self-interest total score</td>
</tr>
<tr>
<td>Other</td>
<td>(SOII) Other-interest total score</td>
</tr>
<tr>
<td>Simple</td>
<td>Target x Instructions</td>
</tr>
</tbody>
</table>