NUSBAUM, EMILY CATHERINE, Ph.D. A Meta-Analysis of Individual Differences in Humor Production and Personality. (2015)
Directed by Dr. Paul J. Silvia. 119 pp.

One main area of focus in humor production research is exploring individual differences in humor production ability (i.e., the ability to produce something funny on the spot), particularly via its relationship with personality. The last 40 years of research, however, has reported conflicting results. Earlier work on individual differences in humor production and personality suggests that extraversion is the most closely related trait to humor production of the Big 5 personality traits. More recent work, however, suggests that openness to experience has the strongest relationship with humor production, and that extraversion has little to no relationship with the ability to produce something funny. The reason for this inconsistency is unclear, but one factor that may contribute to the issue is the between-study variation in assessment of humor production ability and experiment design. One way to resolve this inconsistency is to conduct a research synthesis using meta-analysis, which has two advantages for clarifying the humor production and personality literature: first, it statistically aggregates the findings of completed research in a way that increases statistical power beyond that of the individual studies included in the analysis, and second, it allows for comparison across studies, meaning that random error included in an individual study can be modeled as meaningful variation due to systematic between-study differences. Therefore, the present research meta-analyzed 15 different studies (totaling 56 reported effect sizes) to explore how individual differences in humor production ability relate to personality. Of the Big 5 traits, only openness to experience significantly correlated with humor production ability.

Moderation analyses revealed that while the number of tasks and number of response raters did not have an impact on the size of the openness and humor production effect, the way that humor production ability was modeled did significantly affect the size of the study-level correlation. Finally, moderation analyses revealed that newer assessments of humor production ability did not significantly differ from more traditional assessments. Practical and theoretical implications of these findings for future research are discussed.

A META-ANALYSIS OF INDIVIDUAL DIFFERENCES IN HUMOR PRODUCTION AND PERSONALITY

by

Emily Catherine Nusbaum

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2015

Approved by

_____
Committee Chair

APPROVAL PAGE

This dissertation written by EMILY CATHERINE NUSBAUM has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____

Committee Members _____

_____

_____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

Humor is an important aspect of people's everyday lives. For years researchers have reported correlations between humor and social, emotional, and physical well-being (Boyle & Joss-Reid, 2004; Celso, Ebener, & Burkhead, 2003; Fry, 1994, Kuiper & Nicholl, 2004; Martin, 2004). In particular, humor has been associated with reducing the impact of everyday stressors on health — when people laugh, endorphins are released in the blood stream and positively impact a number of physiological health metrics (Svebak, 2005). Furthermore, humor has been linked to social functioning in both close relationships and more superficial connections (Bressler & Balshine, 2006 Dyck & Holtzman, 2013; Polimeni, Campbell, Gill, Sawatzky, & Reiss, 2010; Sprecher, & Regan, 2002; Stanley, Lohani, & Isaacowitz, 2014). And in these scenarios, humor (or lack thereof) is very salient—some people are funny, and others are very obviously unfunny (Nusbaum & Silvia 2013a, 2013b, 2014) — but the reasons why people differ so much are mysterious. For such an important aspect of our everyday lives, the question of what funny people are like, and how they are different from unfunny people, deserves more attention than it has had in past literature.

In this project, I will explore how personality — specifically, the Big Five traits (and the closely related HEXACO model) — is associated with individual differences in the ability to be funny. Although the literature addressing this question is relatively small,

it is also more diffuse, as demonstrated in the literature review that follows. As a result, the field doesn't have a well-developed sense of direction or focus. Thus, the goal of the present research is to summarize what the existing literature has already demonstrated in terms of the personality and humor production ability relationship, and to suggest concrete directions for future research in the area. An efficient way to do so involves meta-analyzing the results reported in the existing literature. Meta-analysis serves two important purposes: first, it distills an overall summary of the relationship between two variables from the effects reported in the existing work, and second, it allows for comparisons to be made among individual studies — these comparisons help ascertain which aspects of design, sampling, assessment, or other study-specific characteristics are detrimental or beneficial to the effect size obtained in each study. Therefore, this project uses meta-analysis to accomplish the goals of this research — i.e., exploring how personality relates to people's ability to be funny by examining what the present literature tells us about how these individual differences are associated.

What Is Humor?

Part of the confusion surrounding the question of who is funny and who is not is due to the fact that we use one catchall term to describe many different things. "Humor" in its everyday usage can refer to a *sense of humor* — a description of the individual differences in people's sense of the boundary between what is funny and what is distasteful, inappropriate, or nonsensical (Martin, 2003; Martin & Sullivan, 2013); humor can refer to the *use of humor* — how people employ humor in their everyday lives, whether it's for coping, ostracizing, stress management, relationship facilitation, or other

reasons (Abel, 2002; Caird & Martin, 2014); humor might also refer to people's *perception of* humor — people's responsiveness or sensitivity to noticing humor (Papousek, Schulter, Lackner, Samson, & Freudenthaler, 2014; Veatch, 1998); and finally, people use the term humor to describe *humor production*[1], as in people's ability to produce something funny on the spot.

Together, these four aspects of humor (*sense of, uses of*, *perception of*, and *production of*) are important for understanding a more global psychology of humor — which, for something so salient in everyday life, deserves more attention than research has given it in the past. But one aspect in particular, *humor production*, has been especially overlooked. Thus, the focus of this project is the humble *production* aspect of humor.

Research Findings on Humor Production and Personality

Research on humor production is fairly broad. Studies examining humor production have explored its basic relationships with variables like gender (Greengross, Martin, & Miller, 2011; Mickes, Walker, Parris, Mankoff, & Christenfeld, 2012; Robinson & Smith-Lovin, 2001), intelligence (Greengross & Miller, 2011; Howrigan & MacDonald, 2008; Weisfield et al., 2011), other cognitive abilities (Kozbelt & Nishioka, 2010), personality (Greengross et al., 2011; Moran, Rain, Page-Gould, & Mar, 2014), and even terror management (Long & Greenwood, 2013) and pain tolerance (Zweyer, Velker, & Ruch, 2003). But the only common thread among this scattered literature is an exploration of humor production.

Because we don't know much about who is funny (i.e., good at humor production), a logical place to start exploring differences between funny and less-funny people is the correlations among humor production and other individual difference measures. And because personality is so well-validated and reliably measured, the Big 5 traits (Neuroticism, Extraversion, Openness to Experience, Agreeableness, Conscientiousness) are a good way to organize individual differences in humor production. Although what we know about differences in humor production is sparse, there are some natural predictions we might make about who — in terms of personality — is funny.

In particular, people high in openness to experience might be expected to be funnier, given that high openness is associated with greater generalized knowledge and stronger associative ability and verbal fluency. Indeed, Sneed, McCrae, and Funder (1998) found that people perceive others as higher in openness to experience when the observed people were more humorous. But despite our implicit ideas about what funny people are like, measures of openness to experience don't explicitly address humor production. In fact, not one of the few most widely-used measures of openness to experience mentions humor at all (e.g., NEO-PI-R, NEO-FFI-3, Big Five Inventory, HEXACO-100). Scales typically assess appreciation of the arts, curiosity, and unconventional attitudes. Although humor seems like it might fit in with these items, no assessment actually includes it. The International Personality Item Pool—a collaborative pool of personality assessment items—does include a small group of items assessing a "humor" construct directly, but these items aren't excluded by factor analyses from

typical five-factor assessments. And although the items on this small humor scale do hang together well, none directly address humor production specifically—rather, the set of nine items assessing humor/playfulness focus on uses of humor and sense of humor constructs.[2] Thus, while we might naturally expect funny people to be more open, the lack of a humor construct in openness scales makes it seem unlikely that the humor and openness are simply two sides of the same coin.

On the other hand, we tend to perceive "humor" as an inherently social thing. People tend to laugh more in crowds, and it's actually the speakers, rather than the audience, who tend to laugh more frequently (Provine, 1993, 1996; Scott, 2013). People prefer having friends and romantic partners that are funny and that laugh at their jokes (Bressler & Balshine, 2006; Bressler, Martin, & Balshine, 2006; Regan, Levin, Sprecher, Christopher & Cate, 2000; Sprecher & Regan, 2002). And because social activities are the domain of highly extraverted people, it is natural to expect that people who are high in extraversion are also good at being funny. Some research corroborates this assumption: Koppel and Sechrest (1970) and Köhler and Ruch (1996) found a small correlation between extraversion and humor production (*rs* = .19).

Studies that examine humor production and personality often find significant positive correlations between humor production and openness to experience. Greengross, Martin, and Miller (2011), for example, had 400 college students complete the 60-item NEO-FFI-R (Costa & McCrae, 1992) and a version of the cartoon caption task in which people were given 3 captionless cartoons and told to come up with as many captions as they could within 10 minutes. Six judges rated the captions on a 1 to 7 funniness scale.

The three cartoons were reasonably reliable (αs = .69 to .78). People's funniness score was computed as the average of each judge's score for the highest rated caption. Overall, Greengross et al. found that the only Big 5 trait significantly correlated with humor production was openness to experience (r = .26). As an interesting aside, the authors included an assessment of verbal intelligence (a vocabulary test) and found that humor production correlated significantly with verbal ability (r = .39).

Howrigan and MacDonald (2008) likewise found small but positive correlations between personality and humor production. In this study, 185 college students completed measures of personality (50-item IPIP; Goldberg et al., 2006), intelligence (18-item Raven's Advanced Progressive Matrices; Miller & Tal, 2007; Raven, Raven, & Court, 1998), and humor production. To measure humor in this study, the authors asked participants to complete 6 resume tasks and two other novel types of humor production tasks. In one of those tasks, people were told

> For this task, I want you to imagine that you've just received an e-mail by a fellow student asking if you could write some responses to the questions posed below. Your fellow student mentions that this is for a school project on the diversity of humorous responses, and asks that you try to write something funny for each question.
> Question 1: "If you could experience what it's like to be a different kind of animal for a day, what kind would of animal would you not want to be, and why?"
> Question 2: "How would you make a marriage exciting after the first couple of years?"
> Question 3: "What do you think the world will be like in a hundred years?"
> (Howrigan & MacDonald, 2008, p. 656)

The second novel task was specifically aimed at assessing non-verbal humor production, and asked participants to draw "the funniest, most amusing depiction" of 4 different

animals (e.g., monkey, penguin, octopus, giraffe) and 4 different occupations (e.g., politician, professor, body-builder, artist).

Twenty-eight undergraduates rated the humor tasks for funniness on a 1 to 7 scale. Raters were randomly assigned to blocks of participants so that 4 different judges rated each participant's tasks. The three different tasks were reasonably reliable (αs = .63 to .72) and reflected the reliabilities reported in earlier humor production research. Judges' scores were standardized and averaged to produce an overall humor score for each participant. Humor production scores correlated significantly, positively, and equally with openness to experience and extraversion (*r*s = .17). Notably, the authors found that humor production also significantly and positively correlated with performance on the intelligence task (*r* = .29). Taken together, this small literature seems to indicate that the relationship between humor production and personality closely imitates the relationship between openness to experience and creativity, suggesting that the two skills — humor production and creative ideation — may be closely related.

The effects of other personality domains on humor production are less clear. Analyses often fail to identify significant effects for agreeableness, neuroticism, and conscientiousness, and only sporadically identify significant effects for extraversion in either direction. While Babad (1974), Howrigan and MacDonald (2008), and Köhler and Ruch (1996) found significant positive associations between extraversion and humor production (unspecified significant *t*-test; $r = .17$, $p = .021$; and $r = .19$, $p = .047$ respectively), marginally significant negative correlations were found between the two measures in Moran, Rain, Paige-Gould, and Mar (2014) and one dataset in Nusbaum and

Silvia (2013b) (r = -0.14, p = .078 and r = -0.15, p = .070 respectively), and several other studies reported non-significant correlations between the two (Feingold & Mazzella, 1993; Greengross et al., 2011, Kaufman et al., 2013; Koppel & Sechrest, 1970; Nusbaum & Silvia, 2012a, 2012b, 2013b, 2014). Intuitively, the relationship makes sense: extraverted people are affable, energetic, garrulous, and thrive in the spotlight — all qualities that exemplify the stereotypical class clown. But this relationship, especially given its prominence in the humor production literature, deserves a closer examination.

Quantifying Funniness

The inconsistent nature of these findings is perplexing. One likely cause of the variability stems from how humor production is assessed. Most basically, in the study of humor production, participants complete a generation task (i.e., they come up with something funny) and other people rate those responses for funniness. But the variability surrounding this task involves aspects of what sort of task participants do, (i.e., cartoon caption, joke completion, or funny definitions, to name a few) and how those task responses are rated.

Researchers have used humor production tasks that ask people to come up with something funny on the spot by writing a cartoon caption, drawing a funny picture, writing a funny resume, finishing an incomplete joke, or coming up with a funny definition. In the earliest work, Smith and Goodchilds (1963) evaluated the function of "deliberate wits" in group problem-solving activities and recorded instances of joking among group members. But in the 45 years since, researchers have continued using essentially the same measure. Koppel and Sechrest (1970) appear to have originated the

most popular of humor assessment tasks, in which participants are shown a captionless cartoon and instructed to write a funny caption for the cartoon. Because this is a landmark study in the assessment of humor production ability, we should examine it in some detail.

In this experiment, 62 men from fraternities at metropolitan universities were shown a series of 10 cartoons and given one to two minutes to write down the funniest caption that they could think of for each one. The cartoons used in this experiment were single-pane cartoons that came from *The New Yorker* and *Medical Economics* magazines. The selected cartoons were strategically chosen because they depicted a structured scene—that is, they displayed "a relatively complex and suggestive content, such as two people clinging to a plank after a shipwreck rather than two people sitting in chairs" (Koppel & Sechrest, 1970, p. 80). The authors reasoned that these more structured cartoons would make the task less abstract, provide the participants with some direction, and ultimately make it easier for participants to complete the task. Twenty psychology graduate students then rated the captions generated by participants on a 1 to 5 scale of funniness, and the mean of those ratings was the level of each participant's funniness. The task seemed to work well—correlations among self- and peer-rated humor production, and self- and graduate student-rated humor production were sizable (rs = .62 and .43 respectively, $ps < .01$).

A second study published around the same time as Koppel and Sechrest (1970) also employed a cartoon captioning task (Treadwell, 1970). In this study, participants were instructed to write a "humorous and appropriate "caption for each of 11 cartoons that had been drawn specifically for the experiment by a graphic designer colleague of

the author. The captions were then rated on a 1 to 5 funniness scale by two raters. The ratings were then standardized and summed for a final humor score for each participant. In this experiment, humor scores were positively correlated with the ability to solve remote association puzzles (an indicator of creativity) (r = .243), and with the ability to reorganize or redefine concepts (r = .275). The cartoon caption task has endured through 45 years of humor research, with only small changes in cartoon stimuli, number of cartoon captions requested for each cartoon, number of cartoon captioning tasks, time allowed on task, and the scoring procedure for the task.

A few years after these studies, for example, Babad (1974) gave participants 15 minutes to write one funny caption for 15 different captionless cartoons. The cartoon tasks were initially scored three ways: the number of captions people came up with, the number of those captions that were retrieved from memory, and the number of those captions that were original productions (i.e., created during the task). In a second round of scoring, 13 psychology graduate students categorized each caption as either funny or unfunny; if a caption was designated as funny by at least 7 of the 13 judges, it counted toward three new scores: the total number of funny captions, the number of funny captions retrieved from memory, and the number of original funny captions. Although Babad (1974) did not report where the cartoons came from and used a measure of fluency for the humor scores, this study closely resembled the gist of the earlier two studies (Koppel & Sechrest, 1970; Treadwell, 1970). A second study around this same time likewise closely resembled these earlier studies: Brodzinsky and Rubien (1976) had students write funny captions for 6 different single-pane cartoons, five judges rated the

captions on a 1 to 5 funniness scale, and the ratings were averaged across caption tasks and raters to create a mean humor score. In the decade following these studies, the small world of humor researchers continued to rely on these caption tasks to assess humor ability, with varying numbers of tasks, sources of cartoons, and number of raters using a typical 1 to 5 funniness scale. (Masten, 1986; Turner 1980).

After a period of silence in the humor production literature, Feingold and Mazella (1993) returned to the cartoon captions task while developing a multi-dimensional model of wittiness. In one sample, the researchers pulled 8 cartoons from *The New Yorker* magazine, removed the captions, and gave participants unlimited time to write a funny caption for each cartoon. The humor score in this study was computed from two judges' ratings: every caption was scored on a 1 to 5 funniness scale, the scores for each participant's 8 captions were summed, and finally the sums calculated by the two raters were averaged for an overall humor production score.

Köhler and Ruch (1996) used a similar task in their research on humor production. Participants were given 15 captionless cartoons and are asked to write a funny caption for each — however, participants had unlimited time for this task and weren't restricted to writing one caption for each cartoon. Twelve raters used a 9-point scale to rate the wittiness and originality of each caption, and the overall humor score was a mean of the ratings across the task for the 12 raters.

In more recent work, Kozbelt and Nishioka (2010) slightly altered the cartoon caption task. Instead of writing captions for a cartoon, participants were asked to write a funny caption for each of 20 different publicly accessible photos, and had about one

minute to write a caption for each photo. Twelve raters rated the captions on a 1 to 8 scale for funniness, and the scores were collapsed across tasks and raters to compute a total humor score for each participant.

The cartoon captioning task has clearly experienced a long tenure as the go-to humor production assessment, but in recent years, a few different tasks have emerged. One of those tasks involves writing a funny resume for a pictured subject — essentially, a longer and more in-depth version of writing a cartoon caption — where participants describe the hobbies, interests, occupations, life philosophies, and typical days of each target, with the ultimate goal of coming up with something humorous. This resume task has been used in a couple of different studies in recent years. First, Howrigan and MacDonald (2008) explored the utility of the task in a sample of 185 undergraduates. Participants were asked to complete six of these resumes, which were scored on a 1 to 7 humor scale by 28 undergraduate raters. The total humor score was collapsed across ratings and tasks for each participant. The reliability of this task was questionable in this experiment ($\alpha = .58$), but was vastly improved in another study using a similar task. Nusbaum and Silvia (2013b) found in a sample of 168 undergraduates that three resume tasks (differing from Howrigan & MacDonald only in the photos used) rated by 4 raters was quite reliable ($H = .92$). One of the reasons for this difference may lie in the modeling approach — while Howrigan and MacDonald (2008) collapsed scores into averages, Nusbaum and Silva (2013b) used latent variable modeling to estimate a higher-order humor variable. It is worth noting that like the cartoon captioning task, this resume task appears to be an original attempt at assessing a humor production ability trait.

A second task assessing humor production that has recently emerged in the field is a joke completion task. Nusbaum and Silvia (2013a) adapted this task from earlier work in which creativity was assessed with metaphor production (Beaty & Silvia, 2013). In this task, participants are given a scenario and the beginning of a joke, and are asked to complete the joke in a humorous way. Here's an example of what this task looks like:

> Imagine that your friend invites you over and cooks dinner — and the food is totally horrible and disgusting. Later, when describing it to someone else, you say, "Wow, that food was so bad…"
> Please complete the phrase "Wow, that food was so bad…" with something funny.

Those responses are then rated on a 5-point funniness scale by two to four independent raters. Funny responses that participants have given to this task include things like "wow, that food was so bad that it deserves an evil henchman," or "wow, that food was so bad that my taste buds fell out of my mouth and started whimpering." On the other hand, responses that are consistently rated as not funny include things like, "wow, that food was so bad that the dog wouldn't eat it," or "wow, that food was so bad that I threw up." Other joke stems that we have used in this work follow the same format, but ask people to describe the most boring class they've ever taken, or an honest opinion on friend's terrible singing. In four separate studies (Nusbaum & Silvia, 2012a, 2013a, 2013b, 2014), the joke stem tasks were acceptably reliable. Estimates of joke stem reliability at minimum (H = .65) were comparable to those of the classic cartoon captioning task; the majority of estimates, however, indicated greater internal consistency, with estimates as high as H = .84.

A third novel task that we developed to assess humor production ability is a funny definitions task (Nusbaum & Silvia, 2013b). In this task, participants are shown a nonsense noun-noun combination and must come up with a funny definition for that novel compound word:

> A classic form of humor is coming up with funny definitions for things. So, for this next task, you will be given an unusual noun and asked to come up with a funny definition for it, something that most people would find funny or silly. It's fine to be weird, silly, dirty, ironic, bizarre, or whatever, so long as it's funny. For example, you might define "professor" as "someone who talks in someone else's sleep."

In future work researchers could conceivably choose any nonsense noun-noun combination, but we have found success in our work with "cereal bus," "snuggle war," "yoga bank," and "fruit jar." A "cereal bus," for example, was defined by one participant as "a bus made of cereal," while another participant defined "cereal bus" as "the ghetto version of an ice-cream truck."

As in the previously discussed tasks, participants are given unlimited time to come up with one humorous definition for each compound word. The responses are then rated on a 5-point funniness scale by two to four independent raters. As with the joke stems task, the definition task uses the rater's ratings for each task as indicators for a latent variable. In models such as these, reliability estimates have been consistently strong (*Hs* = .81, .80).

But the tasks all have their idiosyncrasies. For example, consider the cartoon caption task. Although many studies assessing humor production use classic cartoon captions tasks with the same gist of assessment (in which participants must write a funny

caption for a captionless single-panel cartoon), there is little standardization across researchers or studies with this task. Some studies ask participants to write multiple captions for one cartoon, while others limit participants to writing one caption for each of many cartoons. Some procedures limit the amount of time participants can work on coming up with a caption, while others allow participants as much time as they want to produce a caption. Some caption tasks ask participants to come up with their own funny caption, while others force participants to complete fill-in-the-blank partially-composed captions. Finally, researchers draw cartoons for their caption tasks from many different (and not always named) sources, and presumably cover many different cartoon styles and topics.

But besides the type of task used to generate funny responses, there are also inconsistencies in the scoring of these humor production tasks — the number of raters scoring participants' responses and how those responses are scored vary from study to study. Although some suggestions have been provided (see Silvia et al., 2008), there is no consensus regarding how many raters are enough for good inter-rater reliability, and how many are too few. A wide range of raters has been employed in studies — from just 2 raters in Feingold and Mazzella (1991, 1993) and Masten (1986), to 6 raters in Greengross and Miller (2011), to 12 raters in Köhler and Ruch (1996) and Kozbelt and Nishioka (2010) to 28 in Howrigan and MacDonald (2008). In addition, it is unclear if researchers should gender-balance the raters to avoid gender biases in the ratings because, as Martin, Puhlik-Doris, Larsen, Gray, and Weir (2003) point out, there appear to be gender differences in sense of humor.

Humor research would also benefit from having a common rating system. Although some studies rate humor responses on some sort of Likert scale assessing funniness (e.g., 1 (*not at all funny*) to 7 (*very funny*)), few studies have described their guidelines that raters may have used to rate the humor responses. Other studies assess humor with fluency, which is simply how many responses people generate. And perhaps as a result, reliability in these tasks has been limited. A typical Cronbach's alpha for a cartoon captions task, for example, is around .60 to .75 (Feingold & Mazzella 1991, 1993; Greengross et al., 2011; Köhler & Ruch, 1996; Kozbelt & Nishioka, 2010; Masten, 1986; Turner, 1980).

Assessing Personality

In the small world of personality and humor production research, two personality models have served as the focus of most of the work. Because the field is small and these models (and their related assessments) play such an integral role, the following section provides a brief overview of the models' development and assessment.

Five-factor model. The classic five-factor model (FFM) had an early start in the empirical study of personality. Researchers approached the issue of outlining a common set of traits with lexical analysis of trait-descriptive adjectives — all nearly 18,000 of them (Allport & Odbert, 1936). Over the years, several attempts were made in the effort to identify a common core of personality traits; the resulting models came up with models ranging in size from just two factors (Eysenck, 1947) all the way to 16 factors (Cattell, 1948).

As researchers began to reexamine the taxonomy of traits, many of their pursuits lead to a model that included five trait dimensions of personality (Digman, 1990; Fiske, 1949; Goldberg, 1981). Some time in the 1970's or 1980's, researchers began to reach a consensus that personality was likely best described along the five dimensions. Four of those dimensions were relatively stable and agreed upon—Neuroticism, Extraversion, Agreeableness, and Conscientiousness—but the fifth factor's interpretation was less clear. It has at times been labeled "culture," "intelligence," and "creativity" (Digman, 1990), and describes someone who is intellectual, yet dreamy and imaginative — two seemingly disparate traits. As a result, different personality assessments sometimes assign different labels for this trait.

The NEO-PI (Costa & McCrae, 1985), which is probably the most widely used measure of personality, calls the fifth factor "Openness to Experience" and assesses six facets of the domain:

- Fantasy: engages in fantasy and daydreams, imaginative, flaky
- Aesthetics: engages with the arts, sensitive to subtle details
- Feelings: emotive, identifies emotions well, perceptive of interpersonal cues
- Actions: impulsive, engages in diverse array of activities and experiences
- Ideas: intellectual, curious, diverse interests
- Values: open to evaluating and adjusting personal values, scrutinizes authority

The introduction of Costa and McCrae's (1985) NEO inventory ignited a major movement in the field of personality research, and in particular, validating factor structure. Many, many studies have tested the validity, reliability, and applicability of the

various versions of the NEO inventory (Costa & McCrae, 1992; McCrae & Costa, 2007; McCrae, Costa, & Martin, 2005).

The NEO scales have been widely used in research on personality and creativity. Naturally, creativity researchers are most interested in the Openness domain. Indeed, most research concerning openness and creativity uses the NEO measures to assess openness. And much of this research finds that Openness is strongly associated with creativity. But this effect is nonetheless curious. If openness and creativity were two sides of the same construct and the consistent relationship between them was simply due to measurement artifact, we would expect to see items like "I enjoy divergent thinking" or "I make an effort to do things differently" in the assessment of openness. Likewise in humor production research, if the openness scales were only correlated with humor production ability simply because the scales subsumed humor production ability, we would expect to see items like "I try to add some humor to whatever I do" or "I like to tease other people out of their gloomy moods." These items actually belong to a small humor subscale of the IPIP (as discussed above), but none of the items appear in the openness scale. In NEO measures of openness, humor production ability — like creativity — is only indirectly and implicitly associated with the domain and its facets. In fact, none of the items on the NEO Openness to Experience scale directly ask about humor. Thus, it is unlikely that the openness and humor production relationship exists because tests and measures of those constructs assess the same things.

HEXACO model. The HEXACO model began as a lexical taxonomy of personality adjectives. Noting several occurrences of a sixth factor emerging in French (Boies, Lee, Ashton, Pascal, & Nicol, 2001), German (Angleitner & Ostendorf, 1989), Korean (Hahn, Lee, & Ashton, 1999), and Dutch (De Raad, 1992) studies, Ashton et al. (2004a) acquired several other datasets of personality adjectives to explore and compare the factor structures within each. They obtained eight different datasets representing seven different languages, and explored the factor structure of each dataset using principal components analyses. The smallest dataset appeared in the Roman sample (285 adjectives), and the largest appeared in the Dutch sample (551 adjectives).

In all eight datasets, analyses suggested six principal components existed. Although the serial order in which the factors emerged varied, similar components were identified in each analysis. The first common component was characterized in the various languages by words like *exuberant*, *social*, and *talkative*; Ashton et al. (2004a) labeled this component *Extraversion* due to its noted similarity with the five-factor model's Extraversion. The second common component was characterized by words like *good-natured*, *gentle*, and *calm*, and was cautiously labeled *Agreeableness*; most adjectives in this factor represented the literal definition of agreeable (i.e., pleasant, cheerful, and tolerant), but some adjectives loading here described characteristics that conventional five-factor models would call Neuroticism (i.e., irritability and emotionality). The third common component included typical descriptions of five-factor model Conscientiousness (i.e., *careful*, *orderly*, *diligent*, *precise*), so Ashton et al. also labeled it *Conscientiousness*. The fourth component was comprised of words that are typical of

Emotional Instability/Neuroticism (i.e., *oversensitive, anxious, emotional,* and *insecure*) and words describing fearlessness (i.e., *courageous, tough,* and *self-assured*); thus this component was cautiously labeled *Emotionality*. Words like *sincere, genuine, honest, arrogant* (low), *greedy* (low), and *cunning* (low), described component five; due to its blended nature, it was labeled *Honesty–Humility*. The final common factor that emerged in all the datasets included words like *intelligent, artistic, bright, creative, progressive,* and *cultured* and was labeled *Intellect/Imagination*.

Because the same six-factor structure appeared in several diverse languages, the authors speculated that perhaps the prior English lexical analyses—on which the Five-Factor Model is based—were too limited to be representative of the language with the largest vocabulary in the world (computing power in the 1980's forced researchers to cluster synonyms to limit the number of variables, and time constraints limited the number of adjectives participants could rate to around 500; Ashton, Lee, & Goldberg, 2004). Thus, a follow-up study was conducted which again used principal components analysis to explore the factor structure of Goldberg's (1982) set of 1,710 English adjectives describing personality characteristics (Ashton et al., 2004b). And again, six components emerged that were similar to the Dutch, Italian, Korean, German, Polish, French, and Hungarian lexicon factor analyses conducted in Ashton et al. (2004a). Furthermore, the six-factor solution had a superior fit and interpretation when compared to a five-factor solution — which was independently interesting, since the five factors that emerged were slightly different from the conventional Five-Factor Model (Ashton et al., 2004b).

The authors then developed a personality inventory from the existing pool of IPIP items to measure the new six-factor model, which they called the HEXACO model. What resulted was the HEXACO Personality Inventory (HEXACO-PI; Lee & Ashton, 2004), which, along with its related metrics — a revised scale (HEXACO-PI-R; Ashton & Lee, 2008), a 60-item HEXACO-60 (Ashton & Lee, 2009), and a 24-item Brief HEXACO Inventory (BHI; de Vries, 2013) — has repeatedly demonstrated convergent validity with other personality scales (Aghababaei, 2012; Ashton & Lee, 2009; Dunlop, Morrison, Koenig, & Silcox, 2012; Gaughan, Miller, & Lynam, 2012; Lee & Ashton, 2013, 2014; Thalmayer, Saucier, & Eigenhuis, 2011; Wasti, Lee, Ashton, & Somer, 2008). The HEXACO model adapts a two-level taxonomy of traits that measures 24 facets across six higher-order domains. Although the key interest of this review focuses on the openness to experience domain, the taxonomy of personality characteristics in the HEXACO model is fairly novel and its departure from traditional five-factor models is significant enough to deserve at least a brief overview.

One of the most substantial differences between five- and six-factor models is found in the agreeableness domain. Although they share a name, the HEXACO domain *Agreeableness* differs significantly from five-factor models. While the positive end of HEXACO *Agreeableness* resembles the five-factor domain with measures of tolerance, gentleness, and friendliness, the negative end of the spectrum is characterized by anger, resentment, and argumentativeness; its facet subscales are *forgivingness, gentleness, flexibility,* and *patience*. Likewise, *Emotionality* differs somewhat from the corresponding NEO domain Neuroticism — they both assesses anxiety proneness, but HEXACO

subscales for this factor (*fearfulness, anxiety, dependence,* and *sentimentality*) mainly assess aspects of social dependence, while Neuroticism measures depression, hostility, and impulsivity. The sixth factor, *Honesty–Humility* captures some aspects of five-factor agreeableness (i.e., morality, honesty) on the positive end, but introduces constructs for conceitedness, arrogance, and greed on the negative pole, which were not accounted for in five-factor models. It has subscales for *sincerity, fairness, greed avoidance,* and *modesty*.

The three remaining HEXACO domains are conceptually very similar to the corresponding NEO-PI-R five-factor measure: the *Extraversion* domain assesses self-esteem, confidence, and sociability with subscales for *social self-esteem, social boldness, sociability* and *liveliness*; the *Conscientiousness* domain assesses orderliness, discipline, and attention to detail with subscales for *organization, diligence, perfectionism,* and *prudence*; and the *Openness to Experience* domain assesses aesthetic appreciation, curiosity, novelty-seeking, and imagination with subscales for *aesthetic appreciation, inquisitiveness, creativity,* and *unconventionality*.

The primary difference between five- and six-factor Openness lies in people's sensitivity to their own feelings and emotions. HEXACO openness is more accurately characterized by engagement in intellectual, aesthetic, and creative pursuits than by emotional introspection, while Five-Factor openness subsumes traits that include emotional intensity and inspection of one's personal values. As a result, perhaps, of its increased specificity, the HEXACO model is gaining popularity among creativity researchers.

The Present Research

The overarching goal of the present research is to bring some clarity and direction to the burgeoning field of humor production research. It has two aims: (1) to summarize and meta-analytically synthesize the available research on humor production and personality, and (2) to explore how between-study assessment and design factors may contribute to observed correlations between personality and humor production. Meta-analysis handles both of these questions well, as it allows for between-study comparisons and generally has more power to detect significant effects when they exist (Cohn & Becker, 2003). While the study, like most meta-analyses, is largely descriptive and exploratory, it is hypothesized that openness to experience will be the only personality trait significantly correlated with humor production ability (i.e., that all other traits will have a non-significant summary correlation). The above review of the existing literature revealed that the effect of openness to experience on humor production was significant more consistently than the significance of other traits' effects. As for the study-level moderators, the small literature leaves an unclear impression of how these factors impact reported effect sizes, but they will be explored nonetheless. It does seem likely, however, that using more precise analysis strategies like structural equation modeling to estimate humor production ability will be associated with stronger study-level correlation estimates.

CHAPTER II

METHOD

Literature Search and Study Selection

Multiple searches were conducted to identify all studies relevant to this meta-analysis. Using PsycINFO, Google Scholar, and Scopus, I searched for several keywords that are likely to be associated with humor and personality: *humor, personality, openness, NEO,* and *humor production*. I also searched the key journals that publish work in humor production and similar areas: *Journal of Personality, Humor: International Journal of Humor Research, Journal of Personality and Social Psychology, Personality and Individual Differences, European Journal of Personality, Journal of Research in Personality, Psychological Assessment,* and *Psychology of Aesthetics, Creativity, and the Arts*, and I examined the reference list of articles collected from the above search. Finally, I contacted researchers active in personality, creativity, and humor research to collect any unpublished datasets they had accumulated that contain the variables of interest. A preliminary search using the strategies outlined above identified 15 studies and 56 effect sizes with strong potential for inclusion in this meta-analysis.

The inclusion criteria for this study were that (1) studies must include a five-factor personality assessment that is one of the following: NEO scales (any version), Big Five Inventory (any version), HEXACO (any version), Eysenck Personality Questionnaire (EPQ; any version) or related Eysenck assessments (e.g., Maudsley

Personality Inventory from Eysenck, 1959), International Personality Item Pool (any version), or the scale must clearly map on to one of the NEO factors; (2) include a direct assessment of humor production that was rated by other people for humorousness of the response; and (3) appear in English. Because this project is focused solely on the ability to produce humor, studies that reported humor based on peer reports, sense of humor, humor comprehension, or self-reported humor were not included in the analyses. In cases where a study reported multiple humor production outcomes for the *same* set of participants — i.e., for two types of humor production tasks (like joke completion and cartoon captions) — a composite effect and variance were computed. Individual study-level factors are also of interest in this project. As part of the analysis, I explored individual effects of task type. Because facets of the Big Five personality traits may provide a more detailed picture of individual differences in humor production, we specifically searched for studies that included facet-level effects, in addition to overall domain effects. Raw data was requested from active researchers in the humor production field.

Coding Procedure

In accordance with the recommendation of the *Cochrane Handbook for Systematic Reviews* (Higgins & Green, 2011), data were extracted from included studies using a pre-specified coding scheme. In addition to basic study demographics (authors, year of publication, location, number of subjects, type of sample, gender of subjects, type of analysis), reports were coded for information pertaining to assessments of humor and personality, and rater information.

Personality assessment type. Included the scale name and number of items for each scale. NEO = 1, Big Five Inventory = 2, International Personality Item Pool = 3, HEXACO = 4, Eysenck Personality Questionnaire = 5, Maudsley Personality Inventory = 6, all others = 7.

Humor production task. Included the number of humor assessments for each type. 1 = captions, 2 = jokes, 3 = resumes, 4 = drawings, 5 = definitions, other = 6.

Rater information. Number of raters, number of male and female raters. Inter-rater reliability. Rating scale. Rater instructions.

Analysis type. Noted whether studies used composite average scores or structural equation models to estimate humor ability.

Studies were dual-coded by two independent raters on the above factors, and no disagreements emerged.

For all included studies, the correlation between trait-level personality and humor ability was recorded, meaning studies could have as many as six recorded effects (one for each trait). But some studies provided additional information beyond the overall effect size. Most studies compiled an average humor score that was used to estimate correlations with personality variables. For example, Howrigan and MacDonald (2008) administered 17 different humor tasks, but used an average score (across all tasks and all raters) to indicate humor production ability. Other studies like Nusbaum and Silvia (2013a, 2013b, 2014) used latent variable models to estimate a latent higher-order humor variable that that was indicated by observed humor ratings and which could be recorded

as specific task-type effects for a more nuanced examination of factors moderating the personality and humor ability effect. Thus, in those studies that supplied the necessary information, separate task-type effects were recorded in addition to an overall trait effect.

Effect Size Calculation

Because all the studies included in this analysis report correlations, the effect sizes in this analysis are also reported as bivariate correlations (*r*), derived from either the reported correlation and sample size, or the reported correlation and standard error. Following the recommended procedures for meta-analysis of *r* (Rosenthal, 1991; Wilson & Dishman, 2015), reported correlations were transformed into Fisher's *z* for summary effect calculation and covariate analysis.

Statistical Analysis

Fixed-effect versus random-effects models. This project has two goals: first, to explore how individual differences in the ability to produce humor relate to individual differences in personality; and second, to examine how sources of within-study variation like task type and number of raters influence those relationships. To address the first goal, I meta-analyzed the reported correlations among humor production ability and each of the Big Five traits. To execute this, five separate meta-analyses — one for each trait — were conducted. Consequently, this method also eliminates any overlap or nesting of effects in the studies included in this meta-analysis, since each subgroup will be analyzed separately (and thus negating any nesting of effects within studies).

Under the fixed effect model, we operate under a couple of assumptions. First, the fixed effect model assumes that all of the effects reported by studies included in the meta-

analysis reflect a single true effect. That is, there may be slight variance in the reported effects, but the variance is not due to true differences in the effects. Because of that assumption, we also have to make a second assumption that any variation in the effect sizes included in the meta-analysis are due to sampling error. For example, a meta-analysis may include a group of studies in which the reported correlations range from r = .08 to .47; in a fixed effect model, we assume that the unknown *true* effect is exactly the same in all of those studies (perhaps r = .25, for instance), and that deviation from that effect (i.e., the range of reported effects in this hypothetical meta-analysis) is simply due to random sampling error.

As a consequence, the summary estimate produced by a fixed effect meta-analysis is assumed to be an estimate of the unknown true effect, and is calculated as the weighted mean of the observed effects.[3] Thus, the null hypothesis for testing the significance of that effect states that the true effect is zero (or one, for a ratio) — that is, there is no treatment effect (or no group difference). Theoretically, the fixed effect model assumes that given an infinite number of studies to include in the meta-analysis, the summary estimate would be the true effect and there would be no variance in that effect.

The random-effects model, on the other hand, does not assume that there is one true effect among studies compiled for meta-analysis. Rather, the random-effects model allows that there may be actual variation in the unknown true effect. In this model, we assume that reported effect sizes in a meta-analysis are estimates of true effect sizes that should be similar but not necessarily exactly the same across studies. Under this assumption, the true effect is conceptualized as a distribution of similar true effects,

rather than a single true effect. The random-effects model thus assumes that the estimates reported in a meta-analysis vary due to both random sampling error *and* systematic error that can be attributed to any number of known or unknown covariates of the effect examined in the meta-analysis. As a consequence, the summary estimate in a random-effects meta-analysis is a weighted mean[4] that estimates the mean of the distribution of true effects. The null hypothesis in a random-effects meta-analysis then is that the *mean* of the distribution of true effects is zero.

Because a random-effects model assumes a distribution of true effects rather than a single true effect, the theoretical implications of the model are slightly different than in the fixed-effect model. In a fixed-effect model, we assume that if we had a single infinitely large sample, our estimate would be the exact unknown true effect because any sampling error would be eliminated. Thus, a fixed-effect meta-analysis combines many smaller-sample studies with the implication that an infinitely large number of studies would theoretically lead to the summary estimate exactly equaling the *unknown true effect*, since sampling error included in each study's estimate would be washed out. In contrast, a single infinitely large sample in a random-effects model would theoretically lead to an exact estimate of only one of the true effects within the distribution of true effects, and an infinitely large number of studies in a random-effects meta-analysis would theoretically lead the summary estimate to exactly equal the *unknown mean of the distribution of true effects*. Although the random effects model seems to be a more conceptually realistic characterization of the humor production and personality relationship, it naturally carries with it a degree of uncertainty around the overall estimate

that accounts for introduction of additional modeled error. Thus, both the fixed-effect and random-effects models will be reported in this analysis.

Assumption of homogeneity. Meta-analyses are performed under the assumption that the effects reported by each study are homogenous (i.e., they're relatively similar and come from a single population). Two statistics are conventionally used to examine whether this assumption holds for a particular meta-analysis: Q and I^2 (Borenstein, Hedges, Higgins, & Rothstein, 2011).

The first statistic, Q, is a standardized measure that examines the ratio of within-study error (the confidence intervals of the estimates) to the variation between the observed correlations (between-study error). It is calculated as a weighted sum of squares of the variation of observed effects in the analysis. For any meta-analysis, the expected value of Q is simply the degrees of freedom within the study—that is, the number of studies in the analysis (k) minus 1. The degree of heterogeneity in a meta-analysis is determined by comparing the expected value of Q (df) to the actual Q; if the calculated Q is greater than the expected value of Q, this indicates that the observed variation of effects is greater than what we would expect given the within-study error. If Q is less than or equal to df, then the observed between-study variation is less than or equal to what we would expect given the within-study error.

One of the benefits of Q is that we can use a significance test (with a chi-square distribution and $k - 1$ degrees of freedom) to examine the degree of the difference between actual and expected Q. However, like any significance test, the power of the Q test is tied to the size of the sample — in this case, meta-analyses with very large or very

small samples will be either overpowered (and thus every test will be significant because of the sheer size of the sample) or underpowered (and thus unlikely to produce any significant result).

The second statistic, I^2, is another measure testing the degree of heterogeneity among effects in a meta-analysis. Essentially, it is a ratio that tells us the proportion of between-study variance that is due to true variation, rather than random error (true variance/total observed variation). The I^2 statistic has two advantages over the Q statistic: first, although it is derived from the Q value, it is calculated to eliminate the dependence of the statistic on the number of effects; second, the fact that it's a proportion makes it intuitively easier to understand and interpret. However, because I^2 is an absolute value (i.e., describes observed effects), it cannot be subjected to significance testing. Nevertheless, I^2 always falls on a 0% to 100% scale, and thus the degree of heterogeneity (i.e., the proportion of variation in effects that is not due to random error) can easily be characterized as low (about 25% or less), moderate (about 50%), or high (about 75% or more).

Publication bias. In peer-reviewed publications, studies reporting significant tests or large effect sizes tend to be published more often than studies reporting non-significant tests or small effect sizes (Borenstein et al., 2011; Dickerson, 2005). This *publication bias* affects meta-analyses' outcomes as well. When compiling studies for a meta-analysis, the most easily obtained studies come in the form of published work; the unpublished work found in theses, dissertations, conference presentations, and the proverbial "file drawer" are frequently more difficult to obtain, and thus are

underrepresented in many meta-analyses. Data in this "grey literature" are much more likely to contain non-significant, small, or perplexing findings that have kept them out of the more easily accessed peer-reviewed journals. Thus, meta-analyses can be impacted by publication bias when they underrepresent effects from grey literature, leading to misleadingly inflated summary estimates.

Although there is not the same pressure in individual differences research to publish only significant results that we might expect in other areas of psychology, for example, publication bias in this meta-analysis will nonetheless be evaluated with several approaches for evaluating bias.

Funnel plots. One approach is to examine a funnel plot (Light & Pillemer, 1984; Light, Singer, & Willett, 1994) of the included effect sizes. In this approach, publication bias is assessed by examining the symmetry of a graph in which the effect sizes (and mean summary estimate) are plotted on the x-axis and the standard errors of each included effect size are plotted on the y-axis. If a meta-analysis isn't affected by publication bias, we would expect a relatively symmetrical distribution of standard errors around the mean estimate. In the presence of publication bias, however, there would be an obvious asymmetry in the distribution of standard errors about the mean.

Begg and Mazumdar rank correlation. While the funnel plot's main appeal is its simple and efficient utility, it is undoubtedly subjective and thus prone to differing interpretations. Because a more objective approach to examining bias in a funnel plot is often desirable in plots that are particularly open to visual interpretation, researchers developed two approaches for quantitatively assessing bias in funnel plots. One approach

developed by Begg and Mazumdar (1994; Begg, 1994) uses rank correlations between standardized effect sizes and their corresponding variances to estimate Kendall's tau (τ). Using this method, if the ranks are independent of one another, the correlation would be near zero. If the correlations are dependent on one another, rather, then Kendall's τ would be significant, and would suggest the presence of publication bias.

Egger's regression intercept. The second approach to quantifying publication bias in a funnel plot involves linear regression. Egger, Davey Smith, Schneider, and Minder (1997) developed a test wherein the study's *z* (calculated as the study's standardized effect divided by standard error) is regressed on the study's precision (the inverse of the standard error), with the intercept β_0 and slope $\beta_1(\text{precision}_i)$. In this approach, any significant deviation of the intercept from zero indicates bias. Thus, using this approach, it's possible to test the null hypothesis that $\beta_0 = 0$, with the consequence that rejection of the null indicates significant bias.

Considerations. There are some considerations that must be made when using either Begg and Mazumdar's (1994) rank correlation test or Egger's regression intercept (Egger et al., 1997). First, in meta-analyses that are relatively small (i.e., under 10 to 25 included studies), or in which there is severe bias, power for detecting bias with these tests is substantially reduced (Begg & Mazumdar, 1994; Sterne, Gavaghan, & Egger, 2000). According to Sterne and Egger (2005), methods such as these for detecting bias should only be used when there exists clear variation in study sample sizes and at least one or more study with a moderate-to-large sample. Although the present study only

borderline meets these suggested parameters, these assessments of publication bias will be included in this analysis in the interest of thoroughness.

Fail-safe N. Besides the funnel plot, one approach researchers use to evaluate publication bias in a meta-analysis is the fail-safe *N*. This "file drawer" approach estimates how many hypothetically non-significant studies it would take to nullify a significant summary effect in a meta-analysis (Orwin, 1983; Rosenthal, 1979). Essentially, how many non-significant studies would a meta-analysis have to have excluded for the significance of the summary effect to be diminished? If the number is quite small, say 4 to 5, then there is significant concern about publication bias impacting the result of the meta-analysis. If the number is quite large, on the other hand (perhaps around 1,000), then there is less concern about the impact of publication bias, since the existence of upwards of 1,000 unpublished studies on the topic is unlikely.

While this method's intuitiveness and ease of use is attractive to researchers conducting meta-analyses, there has been considerable backlash against this method in recent years. Becker (2005) puts it quite bluntly: "Though it is conceptually attractive and relatively simple to use, Rosenthal's fail-safe *N* is prone to misuse and no statistical criterion is available for interpretation of its values. The fail-safe *N* should be abandoned in favor of other more informative analyses" (p. 111). Researchers arguing against the use of the fail-safe *N* mainly cite problems arising from the method's assumption that the average result of all missing or omitted studies is zero. Begg and Berlin (1988) point out that assuming a null result might be biased itself, and that assuming an opposite effect would lead to a much smaller estimate of the number of missing studies it would take to

nullify the obtained effect. In other words, adding a number of negative effects to a positive estimate reduces the estimate to zero much more quickly than adding a number of zero effects. Essentially, Begg and Berlin highlight the fact that assuming the average effect of missing studies is null leads to an inflated estimate of confidence in a meta-analysis' result.

Furthermore, Sutton, Song, Gilbody, and Abrams (2000) point out that it's a weakness that the fail-safe *N* doesn't incorporate the sample size of included studies in its calculation. Not including sample size means that very small (and imprecise) studies are weighted equally with very large (and more precise) studies. Naturally, it's more plausible that a null result truly exists when it's found in a very large, high-powered study than when it's found in a small, underpowered study—weighting both studies equally in the calculation of a fail-safe *N* intuitively feels like willful ignorance of that fact.

Lastly, Iyengar and Greenhouse (1988) pointed out that the fail-safe *N* pools the estimates of standard deviation (or standardized mean differences), which means it assumes the absence of heterogeneity of effects due to any number of factors like sample size, study quality, or true variation in the estimated effect itself. Regardless, though the current study will be relatively small and significant heterogeneity is expected, I will report the fail-safe *N* in the current study in the interest of inclusivity and thoroughness.

Trim-and-fill. In this method, it is assumed that there are a number of relevant studies missing (due to publication bias) from the meta-analysis, and the goal is to attempt to estimate how many studies are missing, and how the summary effect would change if those studies were included. Duval and Tweedie (2000a) developed the trim-

and-fill analysis as an extension of the traditional funnel plot assessment for publication bias, in which asymmetric studies to the specified left or right side of the estimated mean are trimmed from the analysis, then inverted and placed back in the analysis on the opposite side from which they came. The "true" mean and variance are then re-estimated from the filled funnel plot. This method has become widespread in meta-analysis.

For example, imagine a hypothetical funnel plot in which studies' reported effects are plotted against their standard errors: you would expect to see these points falling about the mean of the reported effects in a pine tree shape—those studies with big samples and small standard errors would fall towards the top of the funnel and very close to the mean, while smaller samples (which are naturally prone to more sampling error) with larger standard errors fall towards the middle and bottom of the funnel and farther from the mean. Ideally, these smaller, less-precise studies would fall about evenly on either side of the mean—we expect that sampling error is random error, and thus the chance for reported effects to end up over- or under-estimating the true mean should be about equal. This hypothetical example, however, is rarely how meta-analyses turn out. Rather, funnel plots are often asymmetrical, with a relatively greater number of studies falling on either the right or left side of the mean, and thus show some bias in the studies selected for the analysis. So imagine a new hypothetical funnel plot in which there is some bias toward included studies overestimating the effect. In this case, you would see more dots on the right side of the funnel than the left. Now imagine picking up some of those dots from the right side of the plot and placing them in the same position on the left side of the plot (i.e., same *y*-value, opposite *x*-value). If you were to re-calculate the

summary estimate using the points on this re-organized plot, you would likely find a less-extreme overall effect. In other words, you would have reduced the bias in your meta-analysis and resulting funnel plot by reassigning the asymmetrical points to increase symmetry. The scenario you've just imagined is basically how Duval and Tweedie's (2000a) trim-and-fill method attempts to reduce bias in a meta-analysis.

Sensitivity analysis. Although the trim-and-fill technique can also be considered a sensitivity analysis because it removes potentially influential study effects, one other assessment of the robustness of the summary effect will be used. A one-study-removed analysis examines the robustness of a meta-analytic finding with an iterative method. In this analysis, multiple passes through the analysis are calculated, with one study in the analysis being removed at each pass. Thus, the summary effect is estimated *k* number of times, and the resulting effect of each analysis is the summary estimate excluding that study.

Moderator analyses. One of the main benefits of using meta-analysis to explore the effect of personality on humor production ability lies in its ability to compare effects between studies and explore how individual study-level factors contribute to the reported effect. To make such comparisons and predictions, we can use meta-regression, which uses study-level factors as predictors and reported effect sizes as outcomes. In this analysis, I am particularly interested in exploring how the number of humor production tasks, the type of humor production task, the number of raters, and the type of analysis influences resulting correlations between personality traits and humor production ability. The current literature varies considerably among these dimensions, so understanding how

study design impacts results in this field would provide some much-needed direction for the field of humor production research.

However, because the literature is relatively small, the regression models that can be run are somewhat limited. As in regular regression where it's recommended to have at least 10 subjects per covariate in the regression model, it is likewise recommended that the minimum number of studies per covariate in a meta-regression model is about 10. Since relatively few studies of humor production exist, meta-regression models in this study will be limited generally to an intercept β_0, one covariate β_1, and one outcome. Hence, this study will have two basic meta-regression models examining the impact of the two study-level variables of interest: $\beta_0 + \beta_1(\text{number of tasks}) + e_i = \text{Trait}_i$ and $\beta_0 + \beta_1(\text{number of raters}) + e_i = \text{Trait}_i$. These two models will be used to predict the effect sizes for a single personality trait outcome (i.e., extraversion, for instance). In the case of categorical moderators, I will conduct a random-effects ANOVA to explore the impact of task type and analysis type on the study-level correlation between personality and humor production. Again, because of the small number of studies, I will conduct two separate ANOVAs to investigate one moderator at a time. It should be noted though that meta-regression and meta-ANOVA will only be used in cases where there is significant heterogeneity in the initial meta-analysis (Borenstein et al., 2011), because it is only in those cases that there is variance in the effect sizes that can be potentially explained.

CHAPTER III

RESULTS

Fifteen studies were obtained for this meta-analysis, with an overall sample size of 2,695 people and a total of 57 effect sizes. 40% of the studies included in this meta-analysis came from unpublished datasets. Table 1 shows the entire list of included studies and their reported effects. Because this meta-analysis separately examines correlations between each personality domain and humor production ability, five distinct meta-analyses were conducted—thus, the results for each meta-analysis are discussed separately below.

Openness to Experience

Earlier, I hypothesized that trait openness would have the strongest correlation with humor production ability out of any of the Big 5 (or in the case of the HEXACO, big 6) traits, and that was indeed what this study found. For the openness to experience and humor production meta-analysis, 10 studies satisfied the inclusion criteria, and 10 effects were included in the analysis. Six of these effects were obtained from unpublished datasets. Overall, the 10 studies yielded a total sample of 2,380 participants. Reported correlations in this meta-analysis ranged from r = -0.320 (Greengross et al., 2011a) to r = 0.554 (Nusbaum & Silvia, 2013b).

The overall correlations between openness to experience and humor production in this analysis were 0.233 (0.186, 0.279) for the fixed effects model, and 0.247 (0.151,

0.337) for the random effects model. Table 3 displays the forest plot for this analysis. In all of the forest plots reported here, the black square plots each study's reported correlation, the lines extending from that box indicate the 95% confidence interval around the study's estimate, and the black diamonds at the bottom of the plot mark the fixed- and random-effects models' overall estimate of the effect size.

Heterogeneity and sensitivity. The Q and I^2 statistics (reported in Table 2) indicate significant heterogeneity — that is, true variation in effect size — among the reported effect sizes ($Q = 29.87$, $df = 9$, $p = <.001$). The I^2 value suggests that almost 70% of the variation in reported effect sizes is due to true variation. The robustness of the summary effect was examined in a one-study removed sensitivity analysis. The summary estimate in this analysis was exactly equal to the original summary estimate ($r = 0.247$), indicating that the estimate in this meta-analysis is robust.

Publication bias. Several metrics were used to examine the possibility that the result of the meta-analyses conducted here were influenced by publication bias. The first step that most researchers take when evaluating possible publication bias in a meta-analysis is a visual inspection of the funnel plot. Figure 1 shows the funnel plot for the openness to experience meta-analysis. In this plot, only one effect size falls far outside the funnel. It is important to note, though, that the effect appears at the bottom of the plot, indicating a less precise estimate. Given that the effect size in question had a sample of only 31 people (Greengross et al., 2011), it's to be expected that this effect may be a poor estimate of the mean effect size. Thus, when a trim-and-fill analysis is conducted on the fixed-effect model (Duval & Tweedie, 2000a, 2000b), it is unsurprising that only the one

study (Greengross et al., 2011) is trimmed from the forest plot, and the overall estimate of the mean effect size remains largely unchanged ($r = 0.23$ for observed values versus $r = 0.22$ for adjusted values). In a random-effects model, no studies are trimmed, thus the estimate remains the same. Kendall's tau was non-significant ($\tau = 0.1556$, $z = 0.626$, 2-tailed $p = 0.531$), as was Egger's regression intercept ($\beta_0 = 0.590$, $SE = 1.493$, 2-tailed $p = 0.703$), further suggesting the absence of bias in this meta-analysis. Rosenthal's (1979) fail-safe N was estimated at 200 missing studies, which is relatively a great deal larger than the $k = 10$ studies included in this meta-analysis, suggesting that publication bias is unlikely.

Extraversion

Early theory about the relationship of personality to humor production ability suggested that extraversion has a significant impact on people's ability to be funny (Feingold & Mazzella, 1991, 1993). Later studies, however, reported mixed findings ranging from $r = -0.30$ (Greengross et al., 2012) to $r = 0.19$ (Köhler & Ruch, 1996). Although it theoretically makes sense that being high in extraversion might be associated with humor production ability, the inconsistency in reported correlations suggests that the data might not support the theory. All 15 studies found in the literature search fit the inclusion criteria for the meta-analysis of extraversion and humor production ability. Thus, this meta-analysis was conducted with 15 effect sizes (60% published) and a total sample size of 2,695 people. Reported effects in this analysis ranged from $r = -0.300$ (Greengross et al., 2011) to $r = 0.19$ (Köhler & Ruch, 1996). The overall correlations

between extraversion and humor production in this analysis were -0.007 (-0.051, 0.036) for the fixed effects model, and -0.009 (-0.066, 0.048) for the random effects model.

Heterogeneity and sensitivity. The Q and I^2 statistics (reported in Table 2) do not indicate significant heterogeneity — that is, variation in true effect sizes — among the reported effect sizes ($Q = 20.316$, $df = 14$, $p = .120$). The I^2 value suggests that about 31% of the variation —a small-to-moderate amount, according to conventional metrics — in reported effect sizes is due to true variation. The robustness of the summary effect was examined in a one-study removed sensitivity analysis. The summary estimate in this analysis was exactly equal to the original summary estimate in the random effects model ($r = -0.009$, $p = .761$), indicating that the estimate in this meta-analysis, as in the openness meta-analysis, is robust. Essentially, this analysis demonstrates that while there is a small but non-significant amount of heterogeneity (i.e., variation due to true differences) among the reported effect sizes, the data overall suggest that, on average, there is no relationship between extraversion and the ability to produce something funny.

Publication bias. Overall, tests revealed no evidence of publication bias in this meta-analysis. A visual inspection of the funnel plot for the extraversion and humor production meta-analysis shown in Figure 2 reveals an almost perfect distribution of observed effect sizes about the mean effect. Most studies with more precise estimates (i.e., larger samples, smaller standard errors) fall within the narrow span at the top of the funnel, and all of the less precise estimates fall evenly about the mean as the funnel widens toward the bottom. As might be expected with such a neatly distributed funnel plot, a trim-and-fill analysis had no effect on the overall estimate of the effect of

extraversion on humor production — that is, the analysis did not adjust for bias by trimming any observed effect sizes, and thus the overall effect size estimate remained unchanged (r = -0.009). Kendall's tau was non-significant (τ = -0.105, z = 0.544, 2-tailed p = 0.586), as was Egger's regression intercept (β_0 = -0.282, SE = 0.794, 2-tailed p = 0.728), further suggesting the absence of bias in this meta-analysis. Rosenthal's (1979) fail-safe N was estimated at 0 missing studies, which reflects the null overall estimate. Because the goal of a fail-safe N test is to estimate how many nonsignificant studies would have to be included to nullify the significant summary effect, it doesn't provide any useful information in a study with an already nonsignificant summary effect.

Agreeableness

The lack of significant correlations between agreeableness and humor production ability in any of the previous literature strongly suggests that overall, we should expect a non-significant summary statistic in this meta-analysis. For the agreeableness and humor production meta-analysis, 10 studies satisfied the inclusion criteria, and thus 10 effects were included in the analysis. Six of these effects were obtained from unpublished datasets. Overall, the 10 studies yielded a total sample of 2,380 participants. Reported correlations in this meta-analysis ranged from r = -0.086 (Nusbaum & Silvia, 2012a) to r = 0.160 (Nusbaum & Silvia, 2013b). The overall correlations between agreeableness and humor production in this analysis were -0.005 (-0.053, 0.044) for both the fixed-effect model and the random-effects model. In cases where the fixed- and random-effects models produce identical estimates, it can be interpreted that there is no significant variation in the *true effect*. In other words, if the fixed- and random-effects models

produce identical estimates, there is no true variation between studies — in this analysis, it suggests that the *true effect* of agreeableness on humor production is zero, and that there is no significant variation around that estimate between studies. Table 5 displays the forest plot for this analysis.

Heterogeneity and sensitivity. The Q and I^2 statistics (reported in Table 2) do not indicate significant heterogeneity — that is, true variation in effect size — among the reported effect sizes ($Q = 8.855$, $df = 9$, $p = .451$). The I^2 value suggests that 0% of the variation in reported effect sizes is due to true variation. Essentially, none of the reported effect sizes are significantly different from zero, or from each other. The robustness of the summary effect was examined in a one-study removed sensitivity analysis. The summary estimate in this analysis was exactly equal to the original summary estimate ($r = -0.005$), indicating that the estimate in this meta-analysis is robust.

Publication bias. As in the two previously reported meta-analyses, several metrics were used to examine whether there is evidence of publication bias influencing this analysis. A visual inspection of the agreeableness and humor production funnel plot (shown in Figure 3) reveals that all of the effect sizes included in this meta-analysis fall within the arms of the funnel, indicating an expected distribution of effect sizes about the summary mean effect size. While a trim-and-fill analysis suggested trimming three studies to the right of the mean for a more balanced positive-to-negative distribution, the resulting summary estimate remained nonsignificant and did not differ much from the original estimates — fixed-effect $r = -0.018$ (-0.065, 0.028), random-effects $r = -0.016$ (-0.065, 0.034). Kendall's tau was non-significant ($\tau = 0.244$, $z = 0.984$, 2-tailed $p =$

0.325). Egger's regression intercept, however, was significant ($\beta_0 = 1.491$, *SE* = 0.592, 2-tailed $p = 0.036$). Because of the overall null summary effect, Rosenthal's (1979) fail-safe *N* was not estimated. Although one test did demonstrate bias in the funnel plot for this analysis, the rest did not provide evidence of bias, suggesting that this analysis likely is not biased.

Conscientiousness

Because there is no previous evidence of a significant relationship between conscientiousness and humor production, it was unsurprising that this meta-analysis estimated a nonsignificant summary effect. In the 10 studies that met the inclusion criteria for this analysis, 60% came from unpublished sources. In a total sample of 2,380 people, reported effects ranged from $r = -0.205$ (Nusbaum & Silvia, 2012b) to $r = 0.080$ (Moran et al., 2014). The estimated summary correlation between conscientiousness and humor production in the fixed-effect model was near zero and nonsignificant: $r = -0.007$, $p = .784$ (-0.055, 0.041). The random-effects model produced the exact same estimates, suggesting that there is no variation in the *true effect* of conscientiousness on humor production. Table 6 displays the forest plot for this analysis.

Heterogeneity and sensitivity. As was hinted at by the identical fixed- and random-effects summary estimates, the *Q* and I^2 statistics (reported in Table 2) indicate there is no significant heterogeneity — that is, true variation in effect size — among the reported effect sizes ($Q = 8.448$, $df = 9$, $p = .490$). The I^2 value suggests that 0% of the variation in reported effect sizes is due to true variation. Essentially, none of the reported effect sizes are significantly different from zero, or from each other. The robustness of

the summary effect was examined in a one-study removed sensitivity analysis. The summary estimate in this analysis was exactly equal to the original summary estimate (r = -0.007), indicating that the estimate in this meta-analysis is robust.

Publication bias. Figure 4 displays the funnel plot for the conscientiousness and humor production meta-analysis. A visual inspection of this plot suggests there is no evidence of publication bias in this analysis. All of the included effects fall within the arms of the funnel and are reasonably evenly distributed. Thus unsurprisingly, a trim-and-fill analysis did not recommend trimming any studies, and therefore the summary effect remained unchanged. Kendall's tau was nonsignificant (τ = -0.422, z = 1.699, 2-tailed p = 0.089), as was Egger's regression intercept (β_0 = -1.318, SE = 0.703, 2-tailed p = 0.098), further suggesting the absence of bias in this meta-analysis. Because this analysis estimated a nonsignificant summary effect, Rosenthal's (1979) fail-safe N was not estimated for this analysis.

Neuroticism

Like the agreeableness and conscientiousness analyses above, there is no evidence to suggest that neuroticism should be significantly associated with humor production. This analysis included 11 studies — 55% of which were obtained from unpublished sources — with a total sample of 2,490 people. Reported effects that were included in this meta-analysis ranged from r = -0.149 (Nusbaum & Silvia, 2012b) to r = 0.09 (Greengross et al., 2011). The fixed-effect model produced an overall of r = -0.023 (-0.071, 0.025), p = .351. The random-effects summary estimate was only negligibly different from the fixed effect model, with r = -0.024 (-0.079, 0.031), p = .389.

Heterogeneity and sensitivity. The Q and I^2 statistics (reported in Table 2) do not indicate the presence of significant heterogeneity — that is, true variation in effect size — among the reported effect sizes ($Q = 11.841$, $df = 10$, $p = .296$). The I^2 value suggests that a minimal amount (15.6%) of the variation in reported effect sizes is due to true variation. Essentially, none of the reported effect sizes are significantly different from zero, or from each other. The robustness of the summary effect was examined in a one-study removed sensitivity analysis. The summary estimate in this analysis was exactly equal to the original random-effects summary estimate ($r = -0.024$), indicating that the estimate in this meta-analysis is robust.

Publication bias. As in the previously reported meta-analyses, several metrics were used to examine whether there is evidence of publication bias influencing this analysis. A visual inspection of the neuroticism and humor production funnel plot (shown in Figure 5) reveals that all but one of the effect sizes included in this meta-analysis (Kaufman et al., 2013) fall within the arms of the funnel, indicating a reasonably expected distribution of effect sizes about the summary mean effect size. However, a trim-and-fill analysis suggested trimming three studies to the right of the mean for a more balanced positive-to-negative distribution, and the resulting summary estimate, although the change was minimal, reached significance $r = -0.068$ (-0.111, -0.025), suggesting that it's possible this analysis could be biased toward published data. Kendall's tau was non-significant ($\tau = 0.273$, $z = 1.168$, 2-tailed $p = 0.243$), as was Egger's regression intercept ($\beta_0 = -0.066$, $SE = 1.007$, 2-tailed $p = 0.949$), further suggesting the absence of bias in

this meta-analysis. Because this analysis estimated a nonsignificant summary effect, Rosenthal's (1979) fail-safe *N* was not estimated for this analysis.

Honesty-Humility

Only one study in this meta-analysis reported an effect for HEXACO honesty-humility (Nusbaum & Silvia, 2013b), and thus, the domain effect doesn't benefit from meta-analysis in which heterogeneity, sensitivity, and publication bias can be assessed. Nevertheless, Nusbaum and Silvia (2013b) reported a nonsignificant correlation of r = -0.113 (p = .389) between humor production ability and honesty-humility.

Meta-Regression and Meta-ANOVA

As specified in the Method, only those meta-analyses that produced either a significant summary estimate or significant heterogeneity will be further examined with meta-regression (in the case of continuous covariates) or analysis of variance (in the case of categorical covariates) of the effect size on specified covariates. As a result, only the openness to experience and extraversion meta-analyses are included in this part of the analysis.

All four of the covariates that were coded for — number of tasks, the number of raters judging those tasks, type of humor production task (captions, jokes, and definitions), and type of analysis (average composite humor score versus structural equation modeling) are interesting potential moderators of the overall correlations for two related reasons. One reason is that much of past research in this area has relied on the cartoon-captioning task, which has demonstrated marginal reliability at best in published work. The conventional reliability standards (Kline, 2000) suggest that reliability in the

range of α = .70 to .80 is acceptable, anything between α = .60 to .70 is poor, and anything below α = .60 is unacceptable. Unfortunately, in many studies — for example, Köhler and Ruch (1996; α = .63), Howrigan and MacDonald (2008; α = .58), or Feingold and Mazzella (1993; α= .57) — these reliabilities fall in the poor-to-unacceptable range. Three possible solutions for increasing measurement reliability in future humor production research include increasing the number of tasks that participants complete to a number sufficient for more advanced modeling techniques, using task types that maximize variability between participants, and obtaining better response ratings.

The second reason why these four covariates were chosen is that time is always at a premium in experiments. People lose interest, motivation, and cognitive control over the course of long studies, and may become inattentive to the tasks; long experiments reduce people's interest in consenting to participate in the first place and earn little goodwill from participants; and if the same measurement accuracy can be achieved with just one or two tasks as with 17 tasks (as in Howrigan & MacDonald, 2008), it is an inefficient to include more tasks, types of task, and raters beyond what is psychometrically justified.

All meta-regression models were run within Comprehensive Meta-Analysis 2.0 software using the unrestricted maximum likelihood computational model. This particular computational model was chosen because it runs the regression using the random-effects model. As was discussed in the introduction and revealed in earlier analyses, the random-effects model is more appropriate in this study than the fixed-effect model. Because the models contain only a single covariate, the model estimates will be interpreted using the

Q-test —which tests an omnibus null hypothesis that all model components are zero — and the Z-test, which tests the null hypothesis that a coefficient is zero holding the other model coefficients constant. Since each model is run with only one covariate, the Z-test and Q-test produce equivalent results despite assuming a standard normal distribution and a chi-squared distribution respectively, but they have slightly different advantages. The main advantage of the Z-test is that it reports coefficients for the slope and the intercept, which intuitively gives an idea of how much impact a covariate has on an outcome.

The main reason for choosing the Q-test, on the other hand, is that it uses a sum of squares approach and thus provides an estimate and test of the residual error term implied in the model — in other words, it provides a test of model goodness-of-fit. Having such an estimate would be useful for exploring whether additional heterogeneity remains in the model after the covariate is accounted for. The components of the Q-test model include Q_{model}, which quantifies the dispersion about the mean that is explained by covariates in the model. This component tests the null hypothesis that none of the covariates in the model are related to effect size and is distributed chi-square with df = p number of covariates. $Q_{residual}$ on the other hand quantifies the variation of studies about mean that isn't accounted for by covariates and tests the null hypothesis is that there is no additional variance left in the model with $k - 2$ degrees of freedom. The final component, Q_{total} is the sum of Q_{model} and $Q_{residual}$, which quantifies the dispersion of study effects about the mean effect with $k - 1$ degrees of freedom. Because of the advantages of both the Z-test and Q-test, both will be reported, with a focus on the Z-test for the covariate slope, and

the Q-test for $Q_{residual}$ (i.e., model fit). All regression coefficients are reported as unstandardized values.

To explore whether task type has an impact on the relationship between personality and humor production ability, two moderation models were conducted. Because task type is a categorical variable, the analysis compared the effect of personality on humor production ability across task-type subgroups in two separate analyses (one for openness to experience and another for extraversion). These were run using Comprehensive Meta-Analysis 2.0 as a fully random-effects ANOVA model with between-study variance assumed to be the same across task-type subgroups (and thus pooled across subgroups). The pooled variance estimate was selected for two reasons: one, since the same participants completed all three task types, we should assume the variance between groups would not differ significantly; and two, if groups do not have at least 5 effects in each, it is not possible to calculate separate variance components for each.

Several null hypotheses are tested in this model. A summary correlation is obtained for each subgroup and tested against the null hypothesis that the effect equals zero, and a *Q* statistic examines whether significant heterogeneity exists between subgroups with the null hypothesis that there is not significant between-group variance. In these analyses, the sample of included studies differed from the samples used in all previous analyses and is displayed in Table 8. Because most published work uses a composite humor variable in which each participant's humor score is their average humor score across all tasks and raters for analyses, it's impossible to tease apart the effects of

different task types in these studies. Therefore, the sample for these moderator analyses was limited to studies that either used latent variable modeling (in which an effect could be identified for each task type) or just one type of humor task. In addition, these analyses could only analyze subgroups in which there was more than one effect size reported.

Finally, a similar fully random-effects ANOVA was used to examine the effect of analysis type on the study-level personality and humor production correlation. As was mentioned above, studies included in these meta-analyses differed based on how the humor production variable was estimated for participants. In some of the studies, researchers created a composite score for humor production, in which a participant's humor production score is the average rating across all of their responses and all of the raters. In other studies, a latent humor production score was estimated using structural equation modeling, which is advantageous for handling measurement error in a model (Kline, 2011). Thus, an important question to ask is whether the way that humor production ability is estimated affects the size of the estimated relationship between personality and humor production. In this analysis, the full cohort of studies used for the meta-analyses was included in the ANOVAs. Studies were coded as using either a composite or a latent humor variable, and an ANOVA was computed for the overall openness effect and the overall extraversion effect. As in the task type ANOVA, a summary correlation is obtained for analysis type subgroup and tested against the null hypothesis that the effect equals zero, while a Q statistic examines whether significant heterogeneity exists between subgroups with the null hypothesis that there is not significant between-group variance

Openness to experience.

*Number of tasks***.** In this analysis, I explored how the study-level openness to experience and humor production relationship was influenced by the number of tasks, using the model $\beta_{0i} + \beta_{1i}$(number of tasks) + e_i = openness and humor production r_i. The results of this analysis are shown in Table 9. The slope of the number of tasks in this model was not significant, with $\beta_1 = -0.005$, $p = 0.371$ and 95% confidence interval (-0.015, 0.006), and the Q-test residual term was marginally nonsignificant $Q_{residual}$ = 15.291(8) $p = .054$. Together, these results indicate that there is no effect of the number of tasks on the study-level openness and humor production correlation, and that the heterogeneity between studies is not likely due to between-study differences in the number of tasks that a were included.

Number of raters. The second analysis examined whether the number of task response raters impacted study-level personality estimates. Here, the number of raters predicted the study-level openness and humor production correlation using the model β_{0i} + β_{1i}(number of raters)+ e_i = openness and humor production r_i. The results of this analysis are shown in Table 10. The slope of number of tasks in this model was not significant, with $\beta_1 = -0.005$, $p = 0.433$ and 95% confidence interval (-0.018, 0.008), and the Q-test residual term was marginally non-significant $Q_{residual}$ = 14.139(8) $p = .078$. Together, these results indicate that there is no effect of number of raters on the study-level openness and humor production correlation, and that heterogeneity between studies is not likely due to differences in the number of raters scoring the humor production tasks.

Task type. In this analysis, the effect of task type on the study-level correlation between humor production and openness to experience. All three task types (cartoon captions, joke completions, and funny definitions) were significantly and positively correlated (*rs* = .241, .331, and .223, respectively) with the study-level personality and humor production effects, meaning that each task type predicted larger study-level effects (see Table 11). The *Q*-statistic indicates that there is not significant heterogeneity between task type groups ($Q = 1.188(2)$, $p = .552$), meaning that none of the task subgroups significantly differ from each other.

Type of analysis. Studies in the humor production literature vary as to whether they used a composite (average) humor score for analyses or whether they used structural equation modeling to estimate a higher-order latent humor variable. In this analysis (shown in Table 12), a fully random-effects ANOVA revealed that for the openness to experience meta-analysis, the type of analysis used significantly and positively correlated with the study-level outcome (composite $r = .148$ and SEM $r = .375$). The two types differed from one another, such that studies using a structural equation approach provided significantly larger study-level correlations ($Q = 7.151(1)$, $p = .007$).

Overall, these analyses suggest that neither the number of tasks, the number of raters, or the type of task used to assess humor production are important between study factors. However, it appears that the type of analysis used does significantly impact the study-level openness to experience and humor production correlation.

Extraversion.

Number of tasks. In a fourth analysis, I explored how the study-level effect of extraversion on humor production was influenced by the number of tasks, using the model $\beta_{0i} + \beta_{1i}$(number of tasks) + e_i = extraversion and humor production r_i. The results of this analysis are shown in Table 13. The slope of number of tasks in this model was non-significant, with $\beta_1 = -0.002$, $p = 0.569$ and 95% confidence interval (-0.009, 0.005), and the Q-test residual term was not significant $Q_{residual} = 14.314(13)$, $p = .352$, suggesting that there is no effect of number of tasks on the study-level extraversion and humor production correlation, and that the subgroups do not differ from one another.

Number of raters. This analysis examined whether the number of people rating responses impacted study-level estimates of the personality and humor production relationship. Here, the effect of number of raters predicted the study-level extraversion and humor production correlation using the model $\beta_{0i} + \beta_{1i}$(number of raters)+ e_i = extraversion and humor production r_i. The results of this analysis are shown in Table 14. In this model, the number of raters significantly predicted study-level extraversion and humor production effects, with $Q_{model} = 10.41(1)$, $p = .001$, but the effect was very small. The Z-test revealed a near-zero slope $\beta_1 = 0.009$ (95% CI: 0.004, 0.015), and the Q-test residual term was not significant $Q_{residual} = 9.906(13)$ $p = .702$. Together, these results suggest that there is a very small increase in the estimated correlation between extraversion and humor production as the number of raters in a study goes up, and that after accounting for number of raters, there is no significant heterogeneity among reported effect sizes. In other words, the very small variation that we see in extraversion

effect sizes included in this meta-analysis is entirely due to the number of raters in the study.

Task type. In this analysis, a fully random-effects ANOVA was used to examine the effect of task type on the study-level correlation between humor production and extraversion. Only the funny definitions task was significantly and negatively correlated (r = -0.143) with the study-level personality and humor production effects. These results suggest that while two of the humor production tasks (captions and jokes) have no impact on the study-level extraversion and humor production correlation, the definitions task is actually negatively correlated with that relationship, such that administering a definitions task as an assessment of humor production reduces the study-level correlation between extraversion and humor production (see Table 15). The Q-statistic indicates that there is not significant heterogeneity between task types (Q = 2.815(2), p = .245), meaning that none of the tasks significantly differed from each other.

Type of analysis. In this analysis (shown in Table 16), a fully random-effects ANOVA revealed that for the extraversion meta-analysis, the type of analysis used did not significantly correlate with the study-level outcome (rs = 0.015 and -0.065, respectively) or differ from one another (Q = 1.625(1), p = .202). Overall, the type of modeling approach did not significantly predict differences in study-level correlations.

Facet-Level Analyses

Not enough studies measured or reported facet-level effects for them to be modeled formally in the meta-analysis. A descriptive look, however, suggests that some useful distinctions might appear at the facet level. Kaufman et al. (2013) measured

personality using the Big Five Aspects Scale (DeYoung, Quilty, & Peterson, 2007), which measures two facets of each trait — in the case of openness to experience, the BFAS measures an openness facet and an intellect facet. In this sample, the openness facet was significantly correlated with humor production ($r = 0.224$, $p < .001$), but the intellect facet was not ($r = .066$, $p = .185$). Facet-level personality traits were also measured in one of the datasets in Nusbaum and Silvia (2013b) using the HEXACO-100 inventory. Of the four facets of HEXACO openness — aesthetic appreciation, creativity, inquisitiveness, and unconventionality — only the inquisitiveness and unconventionality facets were significantly (or marginally significantly) correlated with humor production ability ($r = .221$, $p = .054$ and $r = .274$, $p = .035$, respectively).

Finally, the unpublished thesis on which the Howrigan and MacDonald (2008) article is based (Howrigan, 2007) reports facet-level effects of openness to experience (assessed with Goldberg et al.'s, (2006) 100-item IPIP) on humor production. In a sample of 147 college students, Howrigan found that the only significant (or marginally significant) facet-level predictors of humor production ability were the Intellect facet ($r = .18$, $p = .032$) and the Imagination facet ($r = .16$, $p = .058$). The remaining facets (Artistic Interests, Emotionality, Adventurousness, and Liberalism) had nonsignificant correlations with humor production that ranged from $r = .02$ to $r = .10$.

Together, these preliminary results suggest that some aspects of openness to experience, but not others, are primarily driving the correlation between openness and humor production. Although the results are somewhat inconsistent, they do point more toward the quirky, inquisitive, intellectual aspects of openness than toward the aesthetic,

political, and emotional sensitivity aspects as major drivers of the openness and humor production relationship.

CHAPTER IV

DISCUSSION

This study was the first attempt to synthesize the growing literature on humor production ability and its relationship to personality. In sum, the overall meta-analyses turned out as hypothesized. Of the five meta-analyses conducted, openness to experience was the only trait significantly correlated with humor production ability. Although the correlation was small-to-medium by conventional standards (Cohen, 1988), it was in the expected positive direction, and was not unduly influenced by any one study. The openness analysis had the most between-study variation by far, as was found in the significant Q value (indicating significant heterogeneity of effects) and the I^2 value (indicating that the majority of the variation between studies was due to variation in the true effect).

The meta-analyzed effect of extraversion on humor production ability was essentially zero in both the fixed and random effects models. Although some research and theory (e.g., Köhler & Ruch, 1996) suggested that extraversion may be closely tied to the ability to be funny, other work has produced more variation in the effect, with correlations ranging from moderately negative (Greengross et al., 2011) to near zero (Kaufman et al., 2013; Nusbaum & Silvia, 2012a, 2013b). Aggregating the available data revealed an average effect of extraversion on humor production that was not significantly different from zero, suggesting that contrary to intuition, funny people are not particularly

extraverted. In addition, this analysis found little variation between the reported correlations, suggesting that what variation does exist may be due to random error.

The other three traits that were examined in this meta-analysis — agreeableness, conscientiousness, and neuroticism — revealed non-significant and near-zero average effects on humor production ability. There was little in the literature to suggest that these traits may be related to humor production, but one of the advantages of meta-analysis is the ability to detect trends that may not otherwise be observable in small or underpowered individual studies. The analyses of these three traits found miniscule, non-significant correlations with humor production, and zero heterogeneity in the effects reported by individual studies. Essentially, these analyses showed that there is no relationship between humor production ability and agreeableness, conscientiousness, or neuroticism, and that that is very unlikely to change in future research, since all of the zero-effects were homogenously zero.

Effect of Study-Level Moderators

One question this study raises is why there is so much variation in the true effects of openness. Studies included in this analysis varied widely on the number of humor production tasks people completed and the number of judges who rated the funniness of those tasks. But they also varied on the type of task that was used to assess humor production and the method for estimating humor production ability. All of these study-level factors could have some impact on the reliability and validity of estimates of participants' humor ability, so the present research explored whether these differences in

study design might impact the studies' reported correlations, and thus account for some of the heterogeneity of the effects.

Meta-regression was used to examine the impact of two continuous moderators: number of tasks, and number of raters. Because of the small number of studies available for inclusion, each covariate was run in a separate regression model. Thus, two models were run. The model examining the impact of number of tasks on the study-level openness and humor production correlation was not significant — that is, the number of tasks a study used did not have an effect on the size of the correlation. The second model examined the impact of number of raters on the study-level correlation, and found that this covariate also did not have a significant impact on effect size.

The other two covariates, humor production task type and analysis type, were categorical, and thus analyzed with random-effects ANOVA, assuming a common variance between groups. All of the task-types were significantly and positively correlated with the study-level reported effect size, and none of the task-types significantly differed from each other. In other words, all three task types result in a significant study-level correlation between openness to experience and humor production relationship. Furthermore, using structural equation modeling was significantly correlated with larger study-level effects, strongly suggesting that they type of analysis used impacts the resulting correlation.

These same four moderators were examined in the extraversion and humor production meta-analysis. Although the extraversion effect sizes were not significantly heterogeneous, the I^2 value indicated that a small amount of variation between effect

sizes was due to true variation in the effect, and thus warranted an exploration of possible moderating effects. Two meta-regressions (number of tasks and number of raters) and a random-effects ANOVA (type of task) were run to examine the effect of each of the three moderators.

In the model testing the effect of number of tasks, neither the slope of the covariate nor the goodness of fit ($Q_{residual}$) were significant, suggesting that there is no effect of number of tasks on the study-level extraversion effect because there is no significant variation in the effect. In the model exploring the effect of number of raters, there was a significant effect of the covariate, albeit very near zero, indicating that for people high in extraversion, having more raters very slightly increased the size of the study-level effect of extraversion on humor production. In the ANOVA model exploring differences among task-types, the definition task significantly lowered the size of the extraversion effect. However, none of the tasks were significantly different from one another. Thus, the results seem to suggest that perhaps extraverted people don't perform as well as other people on humor production tasks in general. Alternatively, this result could simply be an artifact of a small sample (only two studies used the definitions task) in which there was no relationship between the trait and humor ability. Finally, an ANOVA assessing the impact of analysis type found no significant differences between types, suggesting that the effect of extraversion on humor production is so small that the type of model used in analyses doesn't influence the effect size.

Given the broader picture of results, it seems likely that many of the moderation models did not produce significant effects because the analyses were underpowered.

According to the authors of a popular text on conducting meta-analyses, meta-regressions and ANOVAs tend to be underpowered because of the relatively small sample sizes most meta-analyses produce—in any regular regression model, a sample of 10 to 15 people would be very small, but meta-analyses of 10 to 15 studies are not uncommon (Borenstein et al., 2011).

Overall, however, publication bias was not a factor in these meta-analyses. Only the neuroticism analysis showed some minimal evidence of bias, while the rest of the trait analyses revealed only non-significant measures of bias. The estimated summary effects were all quite robust, as well, as indicated by one-study removed and trim-and-fill analyses. Two factors were likely responsible for the lack of publication bias. One factor is the composition of the included studies themselves—about half of the studies in each analysis were from unpublished sources. The second factor is probably due to the fact that correlational studies in individual differences research typically report the correlations for all the traits that are assessed, including nonsignificant effects for traits of secondary interest.

Practical Implications

The results of these analyses suggest a few different practical implications for future research on humor production. First, this study showed that there is little-to-no variability within trait-specific personality and humor production correlations, with the notable exception of openness to experience. As is the issue with many studies involving small effect sizes, research moving forward should make it a point to include large samples in order to catch between-person variability in humor production. Likewise,

future meta-analyses of personality and humor production that focus on the goal of comparing study-level covariates should be aware that a much larger number of studies (*k*) will be needed to have sufficiently-powered moderator analyses. Although it is not informative for the current research, power was calculated post-hoc for these analyses to inform future researchers endeavoring to repeat similar meta-analyses: assuming a moderate dispersion of effects, 14 degrees of freedom (9 *df* in openness models), and a = .05, power in the openness and extraversion meta-regressions falls below chance levels (.434 and .339, respectively). Clearly, future meta-analyses will need a much larger sample of effect sizes to illuminate study-to-study variation using meta-regressions and analyses of variance.

Second, meta-regressions in this study suggested that there are no significant differences in the impact on study effect size among three humor production tasks (e.g., jokes, captions, and definitions). Although studies in this meta-analysis used some other tasks (i.e., drawings, resumes), there wasn't sufficient data to evaluate differences among these tasks—either the tasks were unique to one study, or the task-specific effect of personality on humor production was not reported. Combined with the significant effect of analysis type on the openness to experience outcome, this issue points to two suggestions for future work: one, use latent variable modeling, or at least report effects of separate task types along with the composite humor score. Two, use a variety of tasks — they all seem to work well, and further analyses on some of the unpublished data included here suggest that using multiple task types and latent variable estimates of higher-order humor production ability leads to more reliable estimates. For example, four

unpublished datasets from Nusbaum and Silvia (2012b, 2013a, 2013b, 2014) used latent variable estimates of humor production and found that reliabilities for a variety of humor tasks ranged from H = .61 to .92, and the reliability of the higher-order latent humor variable ranged from H = .74 to .93 — a considerable improvement over reliabilities reported in earlier work that used an the mean observed humor score in analyses. And third, there's no variability within trait-level analyses for agreeableness, conscientiousness, neuroticism, and arguably extraversion. On top of that, the only trait with significant variability — openness to experience — is probably the broadest of the traits, with facets ranging from constructs like aesthetic appreciation to intellectual curiosity to novelty-seeking. However, none of the published studies included facet level personality measures, so studying facet-level effects is one of the most fruitful directions for future research.

Theoretical Implications

This work also has some theoretical implications for the study of humor production. One of those implications is that humor production ability is probably closely related to creative thinking. If funny things are unexpected and creative (Weems, 2014), and people high in openness are both funnier (as shown in this meta-analysis) and more creative (Feist, 1998), it's possible then that humor production ability is related to creative thinking ability. If that is indeed the case, we could extrapolate from the much larger creativity literature to predict a few more individual differences that might be associated with the ability to be funny on command.

Evidence from the creativity literature suggests that people who are good at coming up with creative ideas score higher on tasks assessing both fluid and crystallized intelligence (Beaty, Silvia, Nusbaum, Jauk, & Benedek, 2014; Nusbaum & Silvia, 2011; Silvia, Beaty, & Nusbaum, 2013). It is thought that while fluid intelligence provides the executive capacity to connect conceptually distant concepts and manipulate their parts into a creative idea, it is crystallized knowledge that provides the raw materials. For example, Beaty and Silvia (2013) found that when writing creative metaphors, participants who scored higher on tasks assessing fluid and crystallized intelligence wrote the most compelling and creative metaphors; however, when asked to write a conventional metaphor the influence of fluid intelligence became non-significant, leaving a moderate effect of crystallized knowledge.

Likewise, people good at coming up with creative ideas are also more open to experience, which is in turn also correlated with those important cognitive abilities. Li et al. (2015) used voxel-based morphometry to examine MRI brain scans of a sample 250 college students who also completed intelligence, creativity, and Big Five personality assessments. Intriguingly, they found that areas of the brain associated with semantic processing, conceptual understanding, making novel associations, and understanding metaphors were correlated with trait creativity, and furthermore, that that correlation was mediated only by openness to experience, suggesting that certain areas of the brain that are advantageous for creative thinking are also significantly correlated with openness to experience.

So with such evidence for the influence of openness to experience on humor production that the present work has provided, is there any reason future research should still investigate the relationship with extraversion? Although this analysis suggests that there isn't an overall effect of extraversion on humor production, there was some variability between reported effects that needs explaining. As noted earlier, very few studies have assessed facet-level personality traits, which may provide a more nuanced picture of how personality traits relate to humor production ability. We discussed hints earlier that some facets of openness to experience may be driving the trait-level correlation, and it's also possible that the same issue exists in the extraversion analyses.

More likely though, the effect of extraversion in humor ability probably lies in the presentation and delivery of humor, rather than the generation or production of humor. An early study, in fact, found a zero-correlation between people nominated by friends, family as humor "producers" and people nominated as "joke tellers." Although personality was only measured with 30 unspecified items from the 69-item Social Introversion scale of the MMPI (Drake, 1946), a one-way ANOVA did reveal that both humor producers and joke tellers were significantly more extraverted than the rest of the sample. Future research could examine whether HEXACO facets like social boldness, social self-esteem, or liveliness are associated with humor delivery, and whether our intuitive perceptions of good deliverers also leads us to perceive them as better producers. Distinguishing between the ability to deliver humor and the ability to produce humor may lead our understanding of the broader humor construct into fertile territory.

More broadly, though, this work implies that the question of how five-factor personality relates to humor production ability is largely answered: the results of this meta-analysis show that only openness to experience is significantly associated with the ability to be funny. This does not imply, however, that research on personality and humor production has divulged all that there is to know. In addition to developing new assessments, the next step in humor production research should move towards theory-building. There is limited evidence regarding the role of openness to experience facets in predictions about humor production, but this direction is perhaps one of the best ways to begin exploring mechanisms of the openness and humor production relationship. For example, openness may facilitate the *desire* to be funny via the unconventionality facet (HEXACO), because being funny simply amuses the quirky nature of high-openness people. Or, openness may facilitate the *ability* to be funny via a link with strong verbal ability. People high in openness are curious, intellectual, and stimulated by novelty — all factors that may contribute to reading a lot and developing a large vocabulary, which in turn may make it easier to create something funny on the spot. In future work, unpacking models of openness to experience will allow researchers to develop a more detailed framework for how and why openness is related to humor production.

Summary and Conclusion

This study was designed to explore the effect of personality on humor production ability with the ultimate goal of synthesizing the existing literature to distill a few notable directions for the field. A series of meta-analyses, meta-regressions, and moderator analyses confirmed what was initially hypothesized: openness to experience was

significantly and positively associated with humor production ability across a variety of study design factors (i.e., task type, number of raters, number of tasks, and type of statistical analysis), and furthermore, it was the only trait significantly correlated with humor production. Its effects, however, were moderately sized at best, suggesting that future research should emphasize obtaining large sample sizes. Nevertheless, the findings presented here will serve as useful benchmarks of effect size for future researchers estimating power and sample size.

A second goal of this work was to explore the between-study differences that moderate study-level effect sizes. While moderator analyses examining the effect of study-level factors were less clear overall, a couple notable results emerged. First, the way that a humor production score is modeled significantly impacts the size of the relationship between personality and humor production — an ANOVA revealed that using structural equation modeling to estimate a latent humor production score (rather than a composite average score) leads to larger study estimates of the relationship, which strongly suggests that future research would be better served by using latent variable methods. A second ANOVA examining the moderating effect of task type on the personality and humor production relationship found that while all of the included task types were significantly and equally associated with a significant openness/humor production correlation at the study level (i.e., none significantly differed from the others), only the definitions task significantly and negatively predicted the extraversion/humor production relationship, suggesting that some tasks seem to perform better than others in different personality groups. Perhaps more notably, however, the finding in the openness

ANOVA highlights the fact that the newer jokes and definitions tasks work as well as the older cartoon captioning task, suggesting that future research can be more enterprising in its assessment of humor production.

In sum, this meta-analysis identified three concrete suggestions for future research: One, researchers should employ a variety of humor production tasks to assess the construct — while the existing tasks seem to work well, there is some evidence that not all tasks work equally well for all traits. Two, this analysis strongly suggests that future analyses would greatly benefit from using more sophisticated statistical techniques than simple averages — estimating a latent humor production ability with structural equation modeling increased the study-level effect sizes, at least in the openness to experience domain. And finally, to accomplish the suggestions laid out above, future research on personality and humor production ability should prioritize obtaining adequately large samples to capture variability in moderate effect sizes and handle larger, more sophisticated models that estimate many more informative parameters.

REFERENCES

Abel, M. H. (2002). Humor, stress, and coping strategies. *Humor: International Journal of Humor Research*, *15*(4), 365-381. doi:10.1515/humr.15.4.365

Aghababaei, N. (2012). Religious, honest and humble: Looking for the religious person within the HEXACO model of personality structure. *Personality and Individual Differences*, *53*(7), 880-883. doi:10.1016/j.paid.2012.07.005

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, *47*(1), i-171. doi:10.1037/h0093360

Amireault, S., Godin, G., & Vézina-Im, L. (2013). Determinants of physical activity maintenance: A systematic review and meta-analyses. *Health Psychology Review*, *7*(1), 55-91. doi:10.1080/17437199.2012.701060

Angleitner, A., & Ostendorf, F. (1989, July). *Personality factors via self and peer-ratings based on a representative sample of German trait descriptive terms*. Paper presented at the First European Congress of Psychology, Amsterdam, the Netherlands.

Ashton, M. C., & Lee, K. (2008). The prediction of Honesty-Humility-related criteria by the HEXACO and Five-Factor Models of personality. *Journal of Research in Personality*, *42*(5), 1216-1228. doi:10.1016/j.jrp.2008.03.006

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*(4), 340-345. doi:10.1080/00223890902935878

Ashton, M. C., Lee, K., & Goldberg, L. R. (2004). A Hierarchical Analysis of 1,710 English Personality-Descriptive Adjectives. *Journal of Personality and Social Psychology*, *87*(5), 707-721. doi:10.1037/0022-3514.87.5.707

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., & ... De Raad, B. (2004a). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, *86*(2), 356-366. doi:10.1037/0022-3514.86.2.356

Babad, E. Y. (1974). A multi-method approach to the assessment of humor. *Journal of Personality, 42*, 618-631.

Banks, G. C., Batchelor, J. H., & McDaniel, M. A. (2010). Smarter people are (a bit) more symmetrical: A meta-analysis of the relationship between intelligence and fluctuating asymmetry. *Intelligence*, *38*(4), 393-401. doi:10.1016/j.intell.2010.04.003

Beaty, R. E., & Silvia, P. J. (2013). Metaphorically speaking: Cognitive abilities and the production of figurative language. *Memory and Cognition*, *41*(2), 255-267. doi:10.3758/s13421-012-0258-5

Beaty, R. E., Silvia, P. J., Nusbaum, E. C., Jauk, E., & Benedeck, M. (2014). The roles of associative and executive processes in creative cognition. *Memory and Cognition, 42*, 1186-1197.

Becker, B. J. (2005) Failsafe *N* or file-drawer number, in publication bias in meta-analysis: Prevention, assessment and adjustments. In H. R. Rothstein, A. J. Sutton and M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 111-125). Chichester, UK: John Wiley & Sons, Ltd. doi: 10.1002/0470870168.ch2

Begg, C.B. (1994). Publication bias. In H.M. Cooper and L.V. Hedges (eds), *The handbook of research synthesis*. New York: Russell Sage Foundation.

Begg, C. B. & Berlin, J. A. (1988). Publication bias: A problem in interpreting medical data (with discussion). *Journal of the Royal Statistical Society, Series A*, *151*, 419-463.

Begg, C. B. & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*, 1088-1101.

Boies, K., Lee, K., Ashton, M. C., Pascal, S., & Nicol, A. M. (2001). The structure of the French personality lexicon. *European Journal of Personality*, *15*(4), 277-295. doi:10.1002/per.41

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

Boyle, G. J., & Joss-Reid, J. M. (2004). Relationship of humor to health: A psychometric investigation. *British Journal of Health Psychology*, *9*(1), 51-66. doi:10.1348/135910704322778722

Bressler, E. R., & Balshine, S. (2006). The influence of humor on desirability. *Evolution and Human Behavior*, *27*(1), 29-39. doi:10.1016/j.evolhumbehav.2005.06.002

Bressler, E. R., Martin, R. A., & Balshine, S. (2006). Production and appreciation of humor as sexually selected traits. *Evolution and Human Behavior*, *27*(2), 121-130. doi:10.1016/j.evolhumbehav.2005.09.001

Brodzinsky, D. M., & Rubien, J. (1976). Humor production as a function of sex of subject, creativity, and cartoon content. *Journal of Consulting and Clinical Psychology*, *44*(4), 597-600. doi:10.1037/0022-006X.44.4.597

Caird, S., & Martin, R. A. (2014). Relationship-focused humor styles and relationship satisfaction in dating couples: A repeated-measures design. *Humor: International Journal of Humor Research*, *27*(2), 227-247. doi:10.1515/humor-2014-0015

Cattell, R. B. (1948). Primary personality factors in the realm of objective tests. *Journal of Personality*, *16,* 459-487. doi:10.1111/j.1467-6494.1948.tb02301.x

Celso, B. G., Ebener, D., & Burkhead, E. (2003). Humor coping, health status, and life satisfaction among older adults residing in assisted living facilities. *Aging & Mental Health*, *7*(6), 438-445. doi:10.1080/1360786031000159469l

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohn, C. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods*, *8*, 243-253.

Costa, P. T., Jr., & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P.T.,Jr. & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual.* Odessa, FL: Psychological Assessment Resources.

de Raad, B. (1992). The replicability of the Big Five personality dimensions in three word-classes of the Dutch language. *European Journal Of Personality*, *6*(1), 15-29. doi:10.1002/per.2410060103

de Vries, R. E. (2013). The 24-item Brief HEXACO Inventory (BHI). *Journal of Research in Personality*, *47*(6), 871-880. doi:10.1016/j.jrp.2013.09.003

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

Dickerson, K. (2005) Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 11-33). Chichester, UK: John Wiley & Sons, Ltd. doi: 10.1002/0470870168.ch2

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41,* 417-440. doi:10.1146/annurev.ps.41.020190.002221

Drake, L. E. (1946). A social I. E. scale for the minnesota multiphasic personality inventory. *Journal of Applied Psychology, 30*, 51–54.

Dubowitz, T., Subramanian, S. V., Acevedo-Garcia, D., Osypuk, T. L., & Peterson, K. E. (2008). Individual and neighborhood differences in diet among low-income

foreign and U.S.-born women. *Women's Health Issues*, *18*(3), 181-190. doi:10.1016/j.whi.2007.11.001

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *108*(19), 7716-7720. doi:10.1073/pnas.1018601108

Dunlop, P. D., Morrison, D. L., Koenig, J., & Silcox, B. (2012). Comparing the Eysenck and HEXACO models of personality in the prediction of adult delinquency. *European Journal of Personality*, *26*(3), 194-202. doi:10.1002/per.824

Duval, S., & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*, 89-98.

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

Dyck, K. H., & Holtzman, S. (2013). Understanding humor styles and well-being: The importance of social relationships and gender. *Personality and Individual Differences*, *55*(1), 53-58. doi:10.1016/j.paid.2013.01.023

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis tested by a simple, graphical test. *British Medical Journal*, *315*, 629-634.

Eysenck, H. J. (1947). *Dimensions of personality*. Oxford England: Kegan Paul.

Eysenck, H. J. (1959). *Manual of the Maudsley Personality Inventory*. London: University of London Press.

Feingold, A., & Mazzella, R. (1991). Psychometric intelligence and verbal humor ability. *Personality And Individual Differences*, *12*(5), 427-435. doi:10.1016/0191-8869(91)90060-O

Feingold, A., & Mazzella, R. (1993). Preliminary validation of a multidimensional model of wittiness. *Journal of Personality*, *61*(3), 439-456. doi:10.1111/j.1467-6494.1993.tb00288.x

Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality and Social Psychology Review*, *2*, 290-309.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology*, *44*(3), 329-344. doi:10.1037/h0057198

Fry, W. F. (1994). The biology of humor. *Humor: International Journal of Humor Research, 7*, 111-125.

Gaughan, E. T., Miller, J. D., & Lynam, D. R. (2012). Examining the utility of general models of personality in the study of psychopathy: A comparison of the HEXACO-PI-R and NEO PI-R. *Journal of Personality Disorders*, *26*(4), 513-523. doi:10.1521/pedi.2012.26.4.513

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of

public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.

Goldberg, L. R. (1981). Unconfounding situational attributions from uncertain, neutral, and ambiguous ones: A psychometric analysis of descriptions of oneself and various types of others. *Journal of Personality and Social Psychology*, *41*(3), 517-552. doi:10.1037/0022-3514.41.3.517

Grant, D. A., & Van Dongen, H. A. (2013). Individual differences in sleep duration and responses to sleep loss. In P. Shaw, M. Tafti, & M. Thorpy (Eds.), *The genetic basis of sleep and sleep disorders* (pp. 189-196). New York, NY, US: Cambridge University Press. doi:10.1017/CBO9781139649469.020

Greengross, G., Martin, R. A., & Miller, G. (2011). Personality traits, intelligence, humor styles, and humor production ability of professional stand-up comedians compared to college students. *Psychology of Aesthetics, Creativity, and the Arts*, *6*, 74-82. doi:10.1037/a0025774

Greengross, G., & Miller, G. (2011). Humor ability reveals intelligence, predicts mating success, and is higher in males. *Intelligence*, *39*, 188-192. doi:10.1016/j.intell.2011.03.006

Hahn, D., Lee, K., & Ashton, M. C. (1999). A factor analysis of the most frequently used Korean personality trait adjectives. *European Journal of Personality*, *13*(4), 261-282.

Higgins, J. P. T & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions* version 5.1.0 [updated March 2011]. The Cochrane Collaboration. Available from www.cochrane-handbook.org.

Holmes, A. J., Lee, P. H., Hollinshead, M. O., Bakst, L., Roffman, J. L., Smoller, J. W., & Buckner, R. L. (2012). Individual differences in amygdala-medial prefrontal anatomy link negative affect, impaired social functioning, and polygenic depression risk. *The Journal of Neuroscience*, *32*, 18087-18100. doi:10.1523/JNEUROSCI.2531-12.2012

Howrigan, D. P. (2007). *Intentional humor as a mental fitness indicator.* California State University, Long Beach. Long Beach, CA.

Howrigan, D. P., & MacDonald, K. B. (2008). Humor as a mental fitness indicator. *Evolutionary Psychology*, *6*(4), 625-666.

Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem. *Statistical Science*, *3*, 109-135.

Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). London:Routledge.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Köhler, G., & Ruch, W. (1996). Sources of variance in current sense of humor inventories: How much substance, how much method variance?. *Humor: International Journal of Humor Research*, *9*, 363-397. doi:10.1515/humr.1996.9.3-4.363

Koppel, M. A., & Sechrest, L. (1970). A multitrait-multimethod matrix analysis of sense of humor. *Educational and Psychological Measurement*, *30*, 77-85. doi:10.1177/001316447003000107

Kozbelt, A., & Nishioka, K. (2010). Humor comprehension, humor production, and insight: An exploratory study. *Humor: International Journal of Humor Research*, *23*(3), 375-401. doi:10.1515/HUMR.2010.017

Kuiper, N. A., & Nicholl, S. (2004). Thoughts of feeling better? Sense of humor and physical health. *Humor: International Journal of Humor Research*, *17*, 37-66. doi:10.1515/humr.2004.007

Lauriola, M., Panno, A., Levin, I. P., & Lejuez, C. W. (2014). Individual differences in risky decision making: A meta‐analysis of sensation seeking and impulsivity with the balloon analogue risk task. *Journal of Behavioral Decision Making*, *27*(1), 20-36. doi:10.1002/bdm.1784

Lee, K. & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39,* 329-358.

Lee, K., & Ashton, M. C. (2013). Prediction of self- and observer report scores on HEXACO-60 and NEO-FFI scales. *Journal of Research in Personality*, *47*(5), 668-675. doi:10.1016/j.jrp.2013.06.002

Lee, K., & Ashton, M. C. (2014). The dark triad, the big five, and the hexaco model. *Personality and Individual Differences*, doi:10.1016/j.paid.2014.01.048

Li, W., Li, X., Huang, L., Kong, X., Yang, W., Wei, D., & ... Liu, J. (2015). Brain structure links trait creativity to openness to experience. *Social Cognitive and Affective Neuroscience*, *10*(2), 191-198. doi:10.1093/scan/nsu041

Light, R.J., & Pillemer, D.B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

Light, R.J., Singer, J.D., & Willett, J.B. (1994). The visual presentation and interpretation of meta-analyses. In M. Cooper & L.V. Hedges (eds), *The handbook of research synthesis.* New York, NY: Russell Sage Foundation.

Linke, A. C., Vicente-Grabovetsky, A., Mitchell, D. J., & Cusack, R. (2011). Encoding strategy accounts for individual differences in change detection measures of VSTM. *Neuropsychologia*, *49*(6), 1476-1486. doi:10.1016/j.neuropsychologia.2010.11.034

Long, C. R., & Greenwood, D. N. (2013). Joking in the face of death: A terror management approach to humor production. *Humor: International Journal of Humor Research*, *26*(4), 493-509. doi:10.1515/humor-2013-0012

Martin, R. (2003). Sense of humor. In S. J. Lopez, C. R. Snyder (Eds.) , *Positive psychological assessment: A handbook of models and measures* (pp. 313-326). Washington, DC, US: American Psychological Association. doi:10.1037/10612-020

Martin, R. A. (2004). Sense of humor and physical health: Theoretical issues, recent findings, and future directions. *Humor: International Journal of Humor Research*, *17*, 1-19. doi:10.1515/humr.2004.005

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality*, *37*(1), 48-75. doi:10.1016/S0092-6566(02)00534-2

Martin, G. N., & Sullivan, E. (2013). Sense of humor across cultures: A comparison of British, Australian and American respondents. *North American Journal of Psychology*, *15*(2), 375-384.

Masten, A. S. (1986). Humor and competence in school-aged children. *Child Development*, *57*(2), 461-473. doi:10.2307/1130601

McCrae, R. R., & Costa Jr, P. T. (2007). Brief versions of the NEO-PI-3. *Journal of Individual Differences*, *28*(3), 116.

McCrae, R. R., Costa, P. r., & Martin, T. A. (2005). The NEO-PI-3: A More Readable Revised NEO Personality Inventory. *Journal of Personality Assessment*, *84*(3), 261-270. doi:10.1207/s15327752jpa8403_05

Mickes, L., Walker, D. E., Parris, J. L., Mankoff, R., & Christenfeld, N. S. (2012). Who's funny: Gender stereotypes, humor production, and memory bias. *Psychonomic Bulletin and Review*, *19*(1), 108-112. doi:10.3758/s13423-011-0161-2

Mikulincer, M., Shaver, P. R., Cooper, M. L., & Larsen, R. J. (2015). *APA handbook of personality and social psychology, volume 4: Personality processes and individual differences*. Washington, DC, US: American Psychological Association. doi:10.1037/14343-000

Miller, G. F., & Tal, I. R. (2007). Schizotypy versus openness and intelligence as predictors of creativity. *Schizophrenia Research*, *93*(1-3), 317-324. doi:10.1016/j.schres.2007.02.007

Moran, J. M., Rain, M., Page-Gould, E., & Mar, R. A. (2014). Do I amuse you? Asymmetric predictors for humor appreciation and humor production. *Journal of Research in Personality*, *49,* 8-13. doi:10.1016/j.jrp.2013.12.002

Nusbaum, E. C. & Silvia, P. J. (2011). Shivers and timbres: Personality and the experience of chills from music. *Social Psychological and Personality Science, 2,* 199-204.

Nusbaum, E. C. & Silvia, P. J. (2012a). *What's so funny? Evaluating humor ability.* Manuscript in preparation.

Nusbaum, E. C. & Silvia, P. J. (2012b). *Ha ha? Assessing humor production ability.* Manuscript in preparation.

Nusbaum, E. C. & Silvia, P. J. (2013a). *Big five personality and humor production ability.* Manuscript in preparation.

Nusbaum, E. C. & Silvia, P. J. (2013b). *Knock, knock: Who's funny? Individual differences in humor production ability.* Manuscript in preparation.

Nusbaum, E. C. & Silvia, P. J. (2014). *Funny ha-ha or funny odd? Individual differences in personality and cognitive abilities associated with humor production ability.* Manuscript in preparation.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analyis. *Journal of Educational Statistics, 8,* 157-159.

Papousek, I., Schulter, G., Lackner, H. K., Samson, A., & Freudenthaler, H. H. (2014). Experimentally observed responses to humor are related to individual differences in emotion perception and regulation in everyday life. *Humor: International Journal of Humor Research*, *27*(2), 271-286. doi:10.1515/humor-2014-0018

Polimeni, J. O., Campbell, D. W., Gill, D., Sawatzky, B. L., & Reiss, J. P. (2010). Diminished humour perception in schizophrenia: Relationship to social and cognitive functioning. *Journal of Psychiatric Research*, *44*(7), 434-440. doi:10.1016/j.jpsychires.2009.10.003

Provine, R. R. (1993). Laughter punctuates speech: Linguistic, social and gender contexts of laughter. *Ethology*, *95*(4), 291-298. doi:10.1111/j.1439-0310.1993.tb00478.x

Provine, R. R. (1996). Laughter. *American Scientist, 84*(1), 38-45.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for raven's progressive matrices and vocabulary scales. section 4, the advanced progressive matrices*. Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

Regan, P. C., Levin, L., Sprecher, S., Christopher, F. S., & Cate, R. (2000). Partner preferences: What characteristics do men and women desire in their short-term sexual and long-term romantic partners?. *Journal of Psychology and Human Sexuality*, *12*(3), 1-21. doi:10.1300/J056v12n03_01

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*(6), 1236-1256. doi:10.1037/0022-3514.84.6.1236

Robinson, D. T., & Smith-Lovin, L. (2001). Getting a laugh: Gender, status, and humor in task discussions. *Social Forces*, *80*(1), 123-158. doi:10.1353/sof.2001.0085

Robles, E., Emery, N. N., Vargas, P. A., Moreno, A., Marshall, B., Grove, R. C., & Zhang, H. (2014). Patterns of responding on a balloon analogue task reveal individual differences in overall risk-taking: Choice between guaranteed and uncertain cash. *Journal of General Psychology*, *141*(3), 207-227. doi:10.1080/00221309.2014.896781

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638– 641.

Rosenthal, R. (1991b). *Meta-analytic procedures for social research* (Vol. 6). Newbury Park: Sage Publications.

Scott, S. (2013). Laughter—The ordinary and the extraordinary. *The Psychologist*, *26*(4), 264-268.

Silvia, P. J., Beaty, R. E., & Nusbaum, E. C. (2013). Verbal fluency and creativity: General and specific contributions of broad retrieval ability (Gr) factors to divergent thinking. *Intelligence*, *41*, 328-340.

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., . . . & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, *2*, 68– 85.

Smith, E. E., & Goodchilds, J. D. (1963). The wit in large and small established groups. *Psychological Reports*, *13*(1), 273-274. doi:10.2466/pr0.1963.13.1.273

Sneed, C. D., McCrae, R. R., & Funder, D. C. (1998). Lay conceptions of the five-factor model and its indicators. *Personality and Social Psychology Bulletin, 24*(2), 115-126. doi:10.1177/0146167298242001

Sprecher, S., & Regan, P. C. (2002). Liking some things (in some people) more than others: Partner preferences in romantic relationships and friendships. *Journal of Social and Personal Relationships*, *19*(4), 463-481. doi:10.1177/0265407502019004048

Stanley, J. T., Lohani, M., & Isaacowitz, D. M. (2014). Age-related differences in judgments of inappropriate behavior are related to humor style preferences. *Psychology and Aging*, *29*(3), 528-541. doi:10.1037/a0036666

Sterne, J. A. C. & Egger, M. (2005) Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton and M. Borenstein (Eds.), *Publication Bias in meta-analysis: Prevention, assessment and adjustments* (pp. 99-110). Chichester, UK: John Wiley & Sons, Ltd. doi: 10.1002/0470870168.ch2

Sterne, J. A. C., Gavaghan, D., & Egger, M. (2000) Publication bias and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology, 53,* 1119-1129.

Sutton, A. J., Song, F., Gilbody, S. M., & Abrams, K. R. (2000). Modeling publication bias in meta-analysis: A review. *Statistical Methods in Medical Research*, *9*, 421-445.

Svebak, S. (2005). Does humor benefit health and well-being? *Journal of the Norwegian Psychological Association*, *42*(10), 909-912.

te Velde, S. J., van der Aa, N., Boomsma, D. I., van Someren, E. W., de Geus, E. C., Brug, J., & Bartels, M. (2013). Genetic and environmental influences on individual differences in sleep duration during adolescence. *Twin Research and Human Genetics*, *16*(6), 1015-1025. doi:10.1017/thg.2013.74

Thalmayer, A., Saucier, G., & Eigenhuis, A. (2011). Comparative validity of Brief to Medium-Length Big Five and Big Six Personality Questionnaires. *Psychological Assessment*, *23*(4), 995-1009. doi:10.1037/a0024165

Turner, R. G. (1980). Self-monitoring and humor production. *Journal of Personality*, *48*(2), 163-172. doi:10.1111/j.1467-6494.1980.tb00825.x

Treadwell, Y. (1970). Humor and creativity. *Psychological Reports*, *26*(1), 55-58. doi:10.2466/pr0.1970.26.1.55

Veatch, T. C. (1998). A theory of humor. *Humor: International Journal of Humor Research*, *11*(2), 161-215. doi:10.1515/humr.1998.11.2.161

Vedel, A. (2014). The Big Five and tertiary academic performance: A systematic review and meta-analysis. *Personality and Individual Differences*, *71,* 66-76. doi:10.1016/j.paid.2014.07.011

Wasti, S., Lee, K., Ashton, M. C., & Somer, O. (2008). Six Turkish personality factors and the HEXACO model of personality structure. *Journal of Cross-Cultural Psychology*, *39*(6), 665-684. doi:10.1177/0022022108323783

Weems, S. (2014). *Ha!: The science of when we laugh and why.* New York, NY: Basic Books.

Weisfield, G. E., Nowak, N. T., Lacas, T., Weisfeld, C. C., Imamoğlu, E. O., Butovskaya, M., & ... Parkhill, M. R. (2011). Do women seek humorousness in men because it signals intelligence? A cross-cultural test. *Humor: International Journal of Humor Research*, *24*(4), 435-462. doi:10.1515/HUMR.2011.025

Wilson, K. E., & Dishman, R. K. (2015). Personality and physical activity: A systematic review and meta-analysis. *Personality and Individual Differences*, *72,* 230-242. doi:10.1016/j.paid.2014.08.023

Zweyer, K., Velker, B., & Ruch, W. (2004). Do cheerfulness, exhilaration, and humor production moderate pain tolerance? A FACS study. *Humor: International Journal of Humor Research*, *17*, 85-119. doi:10.1515/humr.2004.009

APPENDIX A

TABLES

Table 1

Characteristics and Reported Correlations of Included Studies.

Study	**Year**	**n**	**Pers. inv.**	**Pub. status**	**N**	**E**	**O**	**A**	**C**
Feingold	1993	52	Other	P	—	0.030	—	—	—
Feingold	1993	47	Other	P	—	0.070	—	—	—
Feingold	1993	44	Other	P	—	-0.020	—	—	—
Greengross	2011a	31	FFI	P	0.090	-0.300	-0.320	0.130	-0.190
Greengross	2011b	400	FFI	P	-0.090	-0.040	0.260	-0.020	-0.010
Howrigan	2008	185	IPIP	P	-0.040	0.170	0.170	0.100	-0.050
Kaufman	2013	745	FFI	U	0.105	0.016	0.210	-0.080	0.046
Köhler	1996	110	EPQ	P	-0.110	0.190	—	—	—
Koppel	1970	62	MPI	P	—	0.040	—	—	—
Moran	2014	159	BFI	P	0.000	-0.140	0.070	0.030	0.080
Nusbaum	2012a	195	FFI	U	-0.091	0.000	0.222	-0.086	-0.060
Nusbaum	2012b	147	FFI	U	-0.149	-0.034	0.492	0.031	-0.205
Nusbaum	2013a	168	FFI	U	0.046	-0.005	0.554	0.109	-0.092
Nusbaum	2013b	138	HEX	U	-0.047	-0.147	0.346	0.160	0.041
Nusbaum	2014	212	FFI	U	0.067	-0.090	0.332	0.028	0.010
Total	k=15	n= 2,694		60% Pub.					

Note. FFI = Five Factor Inventory (60 items; NEO-FFI 3; McCrae & Costa, 2004). IPIP = International Personality Item Pool (50 items; Goldberg et al., 2006). EPQ = Eysenck Personality Questionnaire – Revised (102 items; Eysenck, Eysenck, & Barrett, 1985). MPI = Maudsley Personality Inventory (80 items; Eysenck, 1959). BFI = Big Five Inventory (44 items; John & Srivastava, 1999). HEX = HEXACO Personality Inventory – Revised (100 items; Lee & Ashton, 2004). P = Published, U = Unpublished.

Table 2

Summary Results of Meta-Analysis for Each Big 5 Trait.

Trait	Total Number of Effects	Mean *r* Fixed (95% CI)	Mean *r* Random (95% CI)	Q	I^2
Neuroticism	11	-0.023 (-0.071, 0.025)	-0.024 (-0.079, 0.031)	11.841	15.6%
Extraversion	15	-0.007 (-0.051, 0.036)	-0.009 (-0.066, 0.048)	20.316	31.1%
Openness	10	0.233 (0.186, 0.279)	0.247 (0.151, 0.337)	29.870	69.9%
Agreeableness	10	-0.005 (-0.053, 0.044)	-0.005 (-0.053, 0.044)	8.855	0.0%
Conscientiousness	10	-0.007 (-0.055, 0.041)	-0.007 (-0.055, 0.041)	8.448	0.0%
Honesty-Humility	1	-0.113 (-0.355, 0.144)	-0.113 (-0.355, 0.144)	—	—

Table 3

Summary of Humor Production and Openness to Experience Meta-Analysis.

Model	Study name	Trait	Statistics for each study				
			Correlation	Lower limit	Upper limit	Z-Value	p-Value
	Greengross (professionals)	O	-0.320	-0.606	0.039	-1.755	0.079
	Greengross (undergrads)	O	0.260	0.166	0.349	5.302	0.000
	Howrigan	O	0.170	0.026	0.307	2.316	0.021
	Kaufman	O	0.210	0.118	0.298	4.430	0.000
	Moran	O	0.070	-0.087	0.223	0.876	0.381
	Nusbaum 2012b	O	0.492	0.304	0.643	4.693	0.000
	Nusbaum 2013b	O	0.346	0.142	0.522	3.242	0.001
	Nusbaum 2014	O	0.332	0.056	0.561	2.344	0.019
	Nusbaum 2012a	O	0.222	0.069	0.365	2.824	0.005
	Nusbaum 2013a	O	0.554	0.313	0.728	4.081	0.000
Fixed			0.233	0.186	0.279	9.473	0.000
Random			0.247	0.151	0.337	4.959	0.000

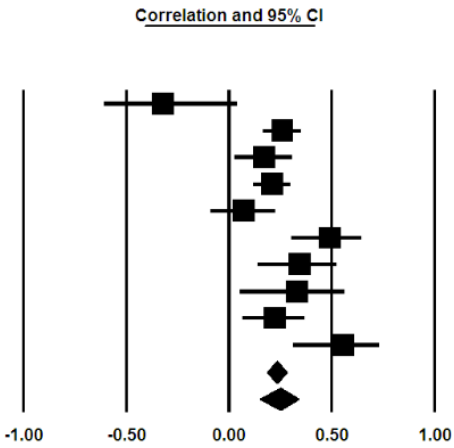


Table 4

Summary of Humor Production and Extraversion Meta-Analysis.

Model	Study name	Trait	Statistics for each study				
			Correlation	Lower limit	Upper limit	Z-Value	p-Value
	Feingold A	E	0.030	-0.245	0.300	0.210	0.834
	Feingold B	E	0.070	-0.222	0.350	0.465	0.642
	Feingold C	E	-0.050	-0.342	0.251	-0.320	0.749
	Greengross (professionals)	E	-0.300	-0.591	0.061	-1.638	0.101
	Greengross (undergrads)	E	-0.040	-0.138	0.058	-0.797	0.425
	Howrigan	E	0.170	0.026	0.307	2.316	0.021
	Kaufman	E	0.016	-0.072	0.104	0.355	0.722
	Kohler	E	0.190	0.003	0.364	1.990	0.047
	Koppel	E	0.040	-0.212	0.287	0.307	0.759
	Moran	E	-0.140	-0.289	0.016	-1.760	0.078
	Nusbaum 2012b	E	-0.034	-0.289	0.225	-0.254	0.800
	Nusbaum 2013b	E	-0.147	-0.299	0.012	-1.811	0.070
	Nusbaum 2014	E	-0.090	-0.286	0.113	-0.869	0.385
	Nusbaum 2012a	E	0.000	-0.190	0.190	0.000	1.000
	Nusbaum 2013a	E	-0.005	-0.221	0.211	-0.045	0.964
Fixed			-0.007	-0.051	0.036	-0.327	0.744
Random			-0.009	-0.066	0.048	-0.304	0.761

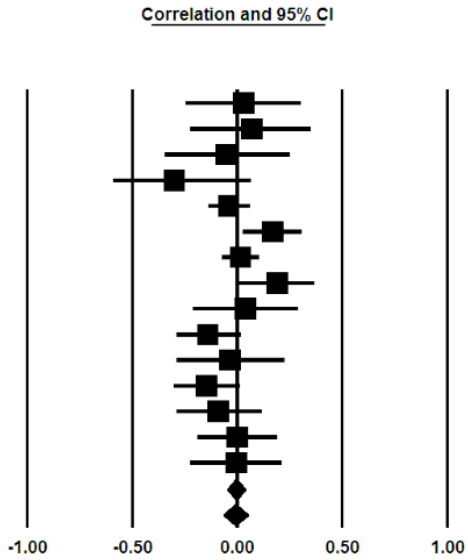


Table 5

Summary of Humor Production and Agreeableness Meta-Analysis.

Model	Study name	Trait	Statistics for each study				
			Correlation	Lower limit	Upper limit	Z-Value	p-Value
	Greengross (professionals)	A	0.130	-0.235	0.463	0.692	0.489
	Greengross (undergrads)	A	-0.020	-0.118	0.078	-0.399	0.690
	Howrigan	A	0.100	-0.045	0.241	1.354	0.176
	Kaufman	A	-0.080	-0.165	0.007	-1.810	0.070
	Moran	A	0.030	-0.126	0.185	0.375	0.708
	Nusbaum 2012b	A	0.031	-0.171	0.231	0.298	0.766
	Nusbaum 2013b	A	0.160	-0.080	0.382	1.310	0.190
	Nusbaum 2014	A	0.028	-0.189	0.243	0.250	0.803
	Nusbaum 2012a	A	-0.086	-0.305	0.142	-0.738	0.461
	Nusbaum 2013a	A	0.109	-0.169	0.371	0.767	0.443
Fixed			-0.005	-0.053	0.044	-0.184	0.854
Random			-0.005	-0.053	0.044	-0.184	0.854

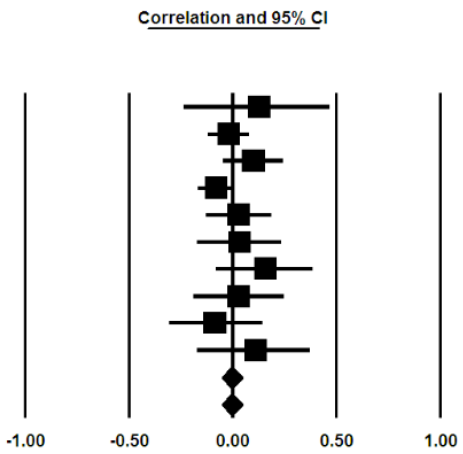


Table 6

Summary of Humor Production and Conscientiousness Meta-Analysis.

Model	Study name	Trait	Statistics for each study				
			Correlation	Lower limit	Upper limit	Z-Value	p-Value
	Greengross (professionals)	C	-0.190	-0.510	0.176	-1.018	0.309
	Greengross (undergrads)	C	-0.010	-0.108	0.088	-0.199	0.842
	Howrigan	C	-0.050	-0.193	0.095	-0.675	0.500
	Kaufman	C	0.046	-0.044	0.136	0.999	0.318
	Moran	C	0.080	-0.077	0.233	1.001	0.317
	Nusbaum 2012b	C	-0.205	-0.391	-0.003	-1.992	0.046
	Nusbaum 2013b	C	0.041	-0.181	0.259	0.359	0.719
	Nusbaum 2014	C	0.010	-0.188	0.207	0.098	0.922
	Nusbaum 2012a	C	-0.060	-0.242	0.126	-0.630	0.529
	Nusbaum 2013a	C	-0.092	-0.337	0.165	-0.698	0.485
Fixed			-0.007	-0.055	0.041	-0.274	0.784
Random			-0.007	-0.055	0.041	-0.274	0.784

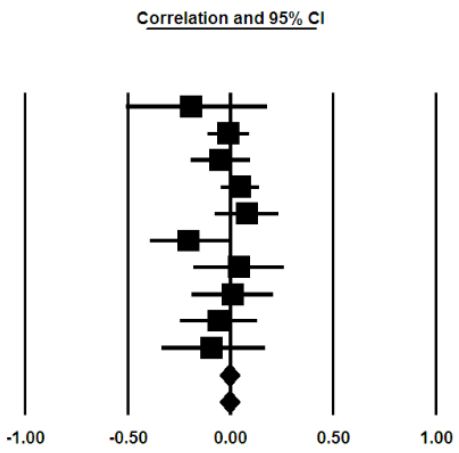


Table 7

Summary of Humor Production and Neuroticism Meta-Analysis.

Model	Study name	Trait	Statistics for each study				
			Correlation	Lower limit	Upper limit	Z-Value	p-Value
	Greengross (professionals)	N	0.090	-0.273	0.431	0.478	0.633
	Greengross (undergrads)	N	-0.090	-0.186	0.008	-1.798	0.072
	Howrigan	N	-0.040	-0.183	0.105	-0.540	0.589
	Kaufman	N	0.105	-0.002	0.209	1.930	0.054
	Kohler	N	-0.110	-0.291	0.079	-1.142	0.253
	Moran	N	0.000	-0.156	0.156	0.000	1.000
	Nusbaum 2012b	N	-0.149	-0.323	0.034	-1.595	0.111
	Nusbaum 2013b	N	-0.047	-0.242	0.152	-0.460	0.645
	Nusbaum 2014	N	0.067	-0.148	0.276	0.607	0.544
	Nusbaum 2012a	N	-0.091	-0.299	0.125	-0.823	0.411
	Nusbaum 2013a	N	0.046	-0.206	0.293	0.353	0.724
Fixed			-0.023	-0.071	0.025	-0.933	0.351
Random			-0.024	-0.079	0.031	-0.862	0.389

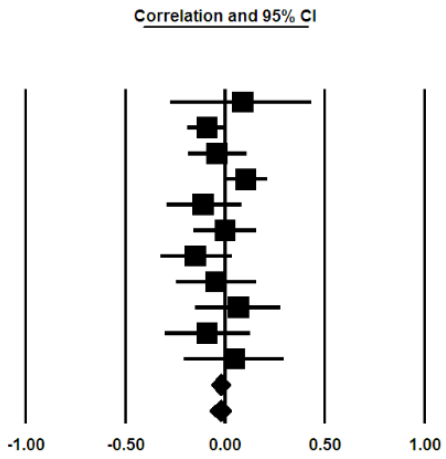


Table 8

Sample of Effects Included in Moderator Analyses.

Study	**Task Type**	**Reported Effect of Openness**	**Reported Effect of Extraversion**
Feingold A	Captions	—	0.030
Feingold B	Captions	—	0.070
Greengross (professionals)	Captions	-0.320	-0.300
Greengross (undergrads)	Captions	0.260**	-0.040
Kaufman 2013	Captions	0.210**	0.016
Nusbaum 2012a	Captions	0.267*	-0.280*
Nusbaum 2013a	Captions	0.213	-0.042
Nusbaum 2013b	Captions	0.558**	0.088
Nusbaum 2014	Captions	0.267**	-0.082
Overall	Captions	0.241**	-0.028
Nusbaum 2012a	Jokes	0.526**	0.036
Nusbaum 2013a	Jokes	0.332**	-0.122
Nusbaum 2013b	Jokes	0.301	-0.208
Nusbaum 2014	Jokes	0.239**	0.049
Overall	Jokes	0.331**	-0.045
Nusbaum 2013b	Definitions	0.226*	-0.187*
Nusbaum 2014	Definitions	0.220**	-0.098
Overall	Definitions	0.223*	-0.143*

*Note***.** * indicates $p < .05$ ** indicates $p < .01$.

Table 9

Meta-Regression of the Effect of Openness to Experience on Number of Humor Production Tasks.

	Estimate	S.E.	95% CI	*p*-Value
β_0	0.296	0.070	(0.159, 0.432)	<0.001
β_1	-0.005	0.005	(-0.015, 0.006)	0.371

	Q	**d.f.**	***p*-Value**
Q_{model}	0.802	1	.371
$Q_{residual}$	15.291	8	.054
Q_{total}	16.093	9	.065

Note. Effects reported here are unstandardized coefficients from the Z-test.

Table 10

Meta-Regression of the Effect of Openness to Experience on Number of Raters.

	Estimate	S.E.	95% CI	*p*-Value
β_0	0.294	0.075	(0.147, 0.441)	<0.001
β_1	-0.005	0.006	(-0.018, 0.008)	0.433

	Q	**d.f.**	***p*-Value**
Q_{model}	0.614	1	.433
$Q_{residual}$	14.139	8	.078
Q_{total}	14.753	9	.098

Note. Effects reported here are unstandardized coefficients from the Z-test.

Table 11

Task Type as a Moderator of the Openness to Experience and Humor Production Correlation.

Task type	***k***	**Estimate**	**95% CI**	***p*-Value**	***Q***	**d.f. (*Q*)**	***p*-Value(*Q*)**
Captions	7	0.241	(0.136, 0.341)	<.001			
Jokes	4	0.331	(0.036, 0.394)	0.020			
Definitions	2	0.223	(0.180, 0.467)	<.001			
Overall	13	0.262	(0.180, 0.341)	<.001			
					1.188	2	0.552

Note. Effects reported here are from the random-effects model.

Table 12

Analysis Type as a Moderator of the Openness and Humor Production Correlation.

Task type	***k***	**Estimate**	**95% CI**	***p*-Value**	***Q***	**d.f. (*Q*)**	***p*-Value(*Q*)**
Average	5	0.148	(0.034, 0.258)	0.011			
SEM	5	0.375	(0.251, 0.487)	<.001			
Overall	10	0.262	(0.029, 0.469)	0.028			
					7.151	1	0.007

Note. Effects reported here are from the random-effects model.

Table 13

Meta-Regression of the Effect of Extraversion on Number of Humor Production Tasks.

	Estimate	S.E.	95% CI	*p*-Value
β_0	0.010	0.043	(-0.074, 0.094)	0.822
β_1	-0.002	0.003	(-0.009, 0.005)	0.569

	Q	**d.f.**	***p*-Value**
Q_{model}	0.325	1	.569
$Q_{residual}$	14.314	13	.352
Q_{total}	14.638	14	.403

Note. Effects reported here are unstandardized coefficients from the Z-test.

Table 14

Meta-Regression of the Effect of Extraversion on Number of Raters.

	Estimate	**S.E.**	**95% CI**	***p*-Value**
β_0	-0.096	0.035	(-0.165, -0.027)	0.007
β_1	0.009	0.003	(0.004, 0.015)	0.001

	Q	**d.f.**	***p*-Value**
Q_{model}	10.410	1	.001
$Q_{residual}$	9.906	13	.702
Q_{total}	20.316	14	.120

Note. Effects reported here are unstandardized coefficients from the Z-test.

Table 15

Task Type as a Moderator of the Extraversion and Humor Production Correlation.

Task type	***k***	**Estimate**	**95% CI**	***p*-Value**	***Q***	**d.f. (*Q*)**	***p*-Value(*Q*)**
Captions	9	-0.028	(-0.083, 0.028)	0.332			
Jokes	4	-0.045	(-0.163, 0.074)	0.460			
Definitions	2	-0.143	(-0.261, -0.020)	0.023			
Overall	15	-0.061	(-0.136, 0.016)	0.121			
					2.815	2	0.245

Note. Effects reported here are from the random-effects model.

Table 16

Analysis Type as a Moderator of the Extraversion and Humor Production Correlation.

Task type	***k***	**Estimate**	**95% CI**	***p*-Value**	***Q***	**d.f. (*Q*)**	***p*-Value(*Q*)**
Average	10	0.015	(-0.052, 0.081)	0.664			
SEM	5	-0.065	(-0.166, 0.038)	0.216			
Overall	15	-0.015	(-0.090, 0.060)	0.697			
					1.625	1	0.202

Note. Effects reported here are from the random-effects model.

APPENDIX B

FIGURES

Figure 1

Funnel Plot of Standard Error by Fisher's Z for the Humor Production and Openness to Experience Meta-Analysis.

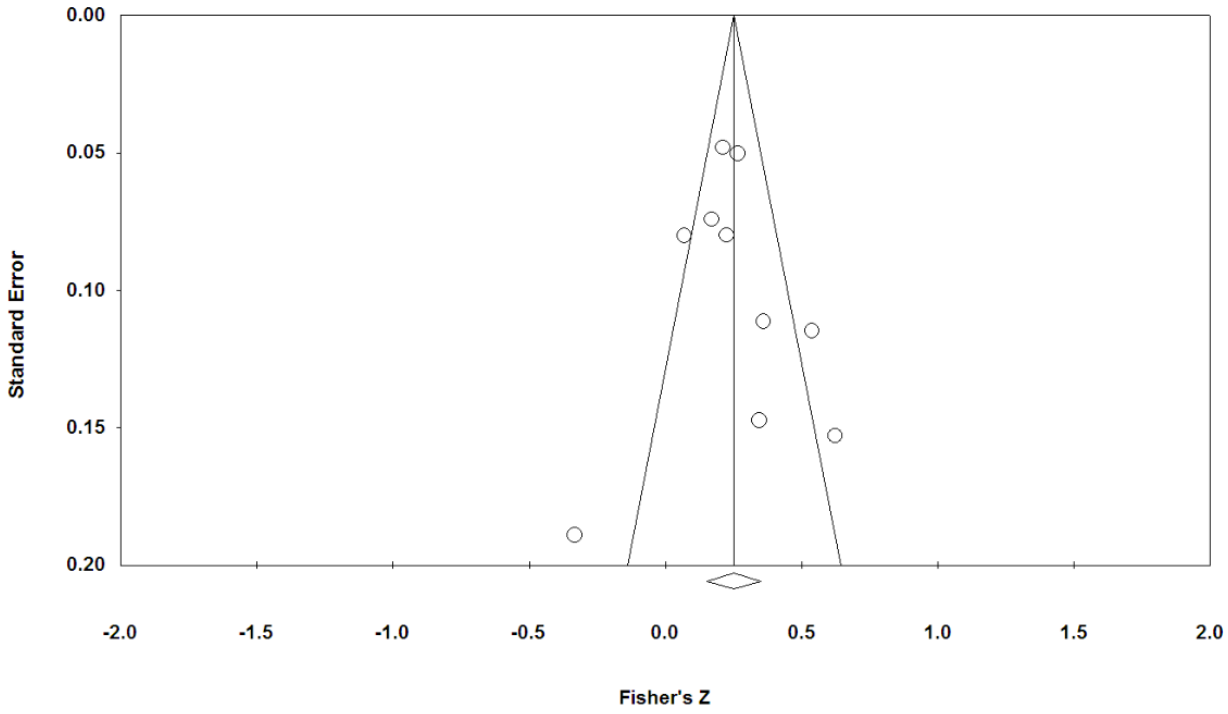


Figure 2

Funnel Plot of Standard Error by Fisher's Z for the Humor Production and Extraversion Meta-Analysis.

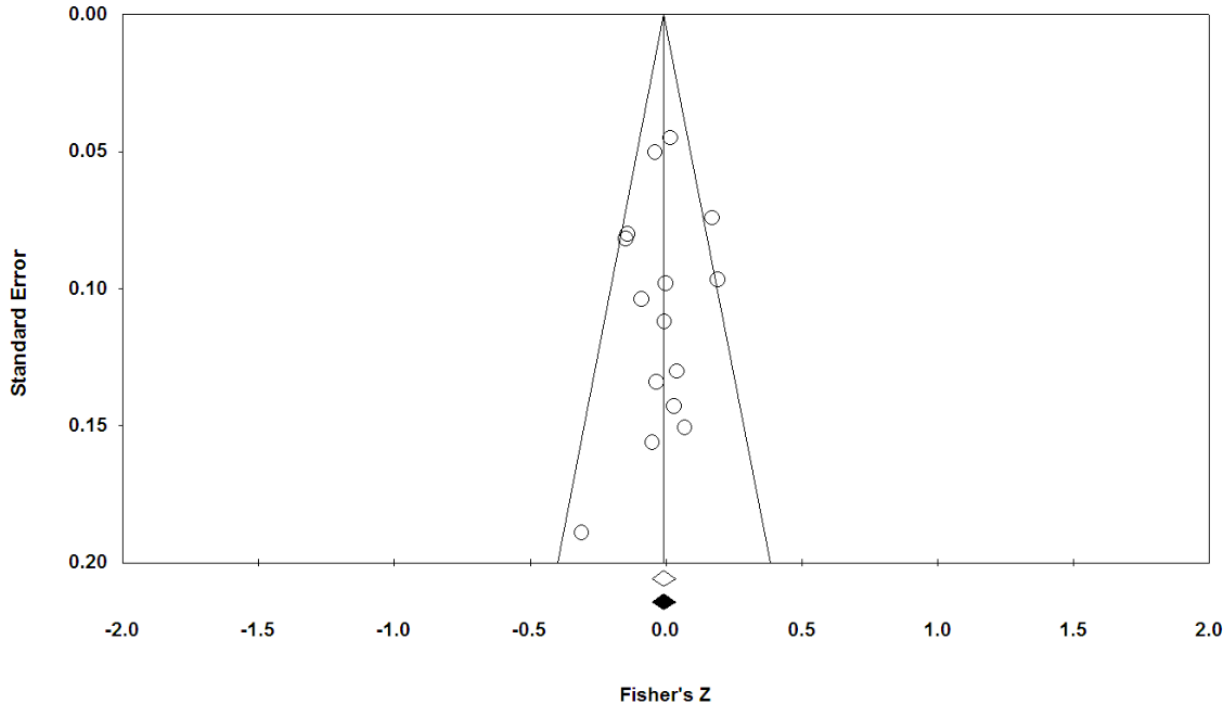


Figure 3

Funnel Plot of Standard Error by Fisher's Z for the Humor Production and Agreeableness Meta-Analysis.

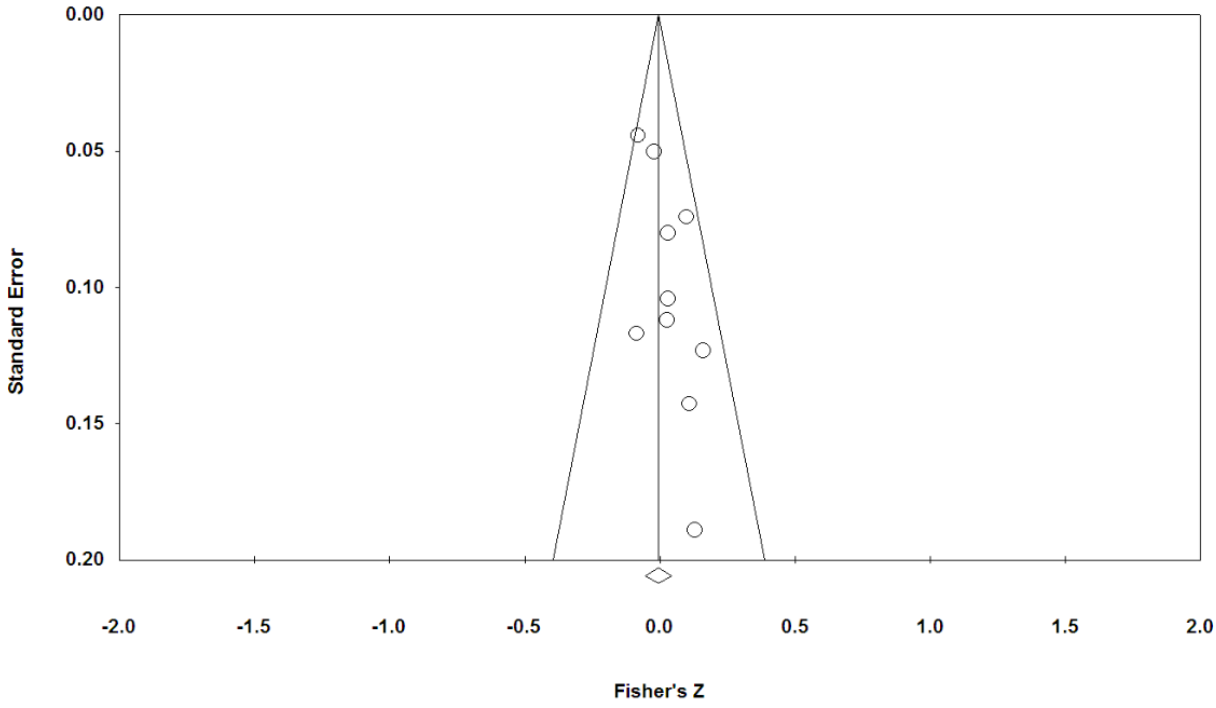


Figure 4

Funnel Plot of Standard Error by Fisher's Z for the Humor Production and Conscientiousness Meta-Analysis.

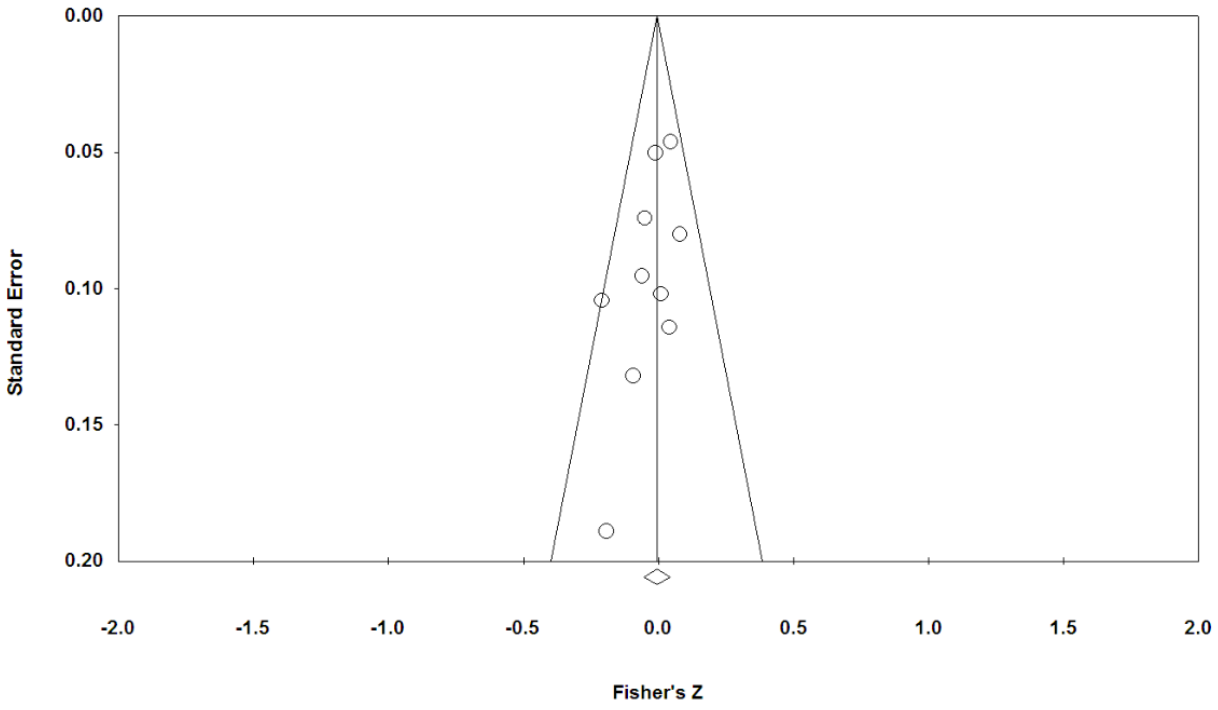
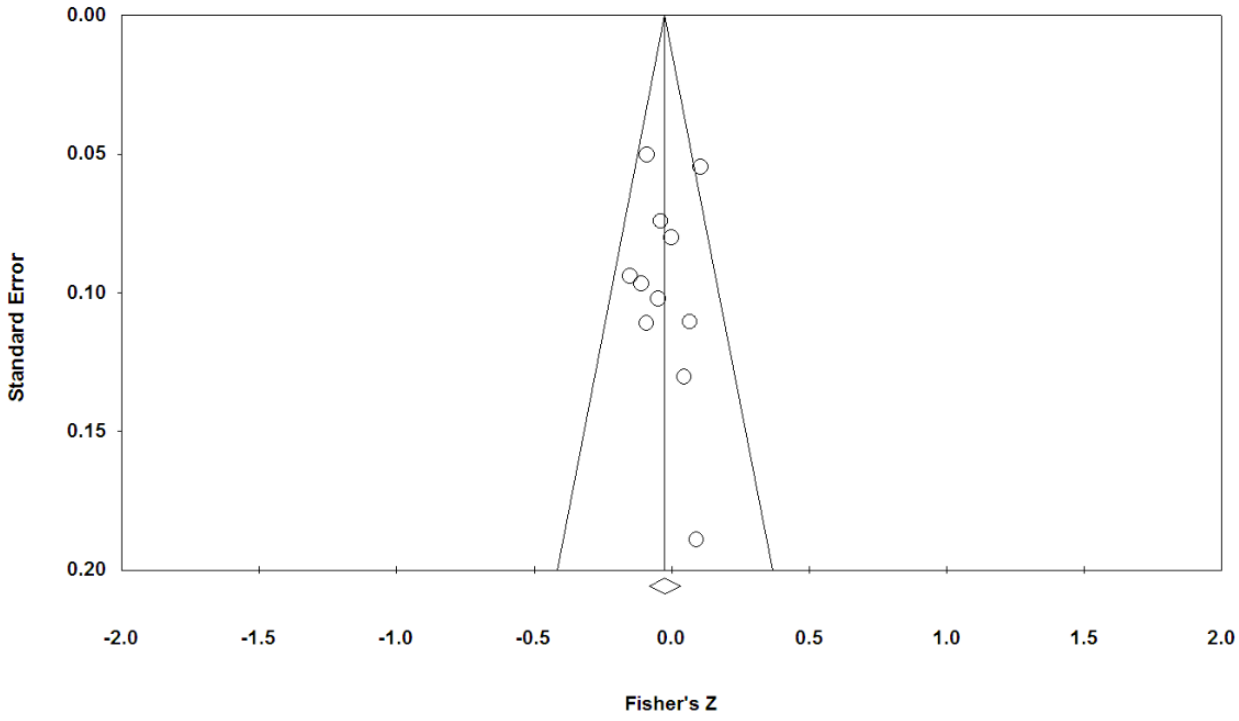


Figure 5

Funnel Plot of Standard Error by Fisher's Z for the Humor Production and Neuroticism Meta-Analysis.



APPENDIX C

ENDNOTES

[1] Readers should note that this construct appears throughout the literature as "humor ability," and the terms appear to be used interchangeably. In the interest of clarity, however, the ability to be funny on the spot will be referred to as "humor production" throughout this paper.

[2] The nine IPIP items on this humor/playfulness scale are: (1) *Try to tease my friends out of their gloomy moods*; (2) *Use laughter to brighten the days of others*; (3) *Try to have fun in all kinds of situations*; (4) *Try to add some humor to whatever I do*; (5) *Keep my sense of humor even in gloomy situations*; (6) *Have a great sense of humor*; (7R) *Am not known for my sense of humor*; (8R) *Am not fun to be with*; (9R) *Do not go out of my way to make others smile or laugh*.

[3] Observed effects in a fixed-effect model are weighted in the summary (mean) effect by the inverse of the their variances.

[4] Observed effects in a random-effects model are weighted in the summary (mean) effect by the inverse of their variances, which, in this model, are calculated to include both within-study and estimated between-study variance.