

McGee, Andrew; McGee, Peter

Working Paper

Whoever You Want Me to Be: Personality and Incentives

IZA Discussion Papers, No. 13809

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: McGee, Andrew; McGee, Peter (2020) : Whoever You Want Me to Be: Personality and Incentives, IZA Discussion Papers, No. 13809, Institute of Labor Economics (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/227336>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

DISCUSSION PAPER SERIES

IZA DP No. 13809

**Whoever You Want Me to Be:
Personality and Incentives**

Andrew McGee
Peter McGee

OCTOBER 2020

DISCUSSION PAPER SERIES

IZA DP No. 13809

Whoever You Want Me to Be: Personality and Incentives

Andrew McGee

University of Alberta and IZA

Peter McGee

University of Arkansas

OCTOBER 2020

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Whoever You Want Me to Be: Personality and Incentives*

What can employers learn from personality tests when job applicants have incentives to misrepresent themselves? Using a within-subject, laboratory experiment, we compare personality measures with and without incentives for misrepresentation. Incentivized personality measures are weakly to moderately correlated with non-incentivized measures in most treatments but are correlated with intelligence when test-takers have information about desired personalities or are warned that responses may be verified. We document that actual job ads provide information about desired personalities and that employers in the UK who administer personality tests are also likely to administer intelligence tests despite the potential for substitution between the tests.

JEL Classification: C91, D82, M50

Keywords: personality, measurement, hiring, screening, experiments

Corresponding author:

Andrew McGee
Department of Economics
University of Alberta
8-14 Tory Building
Edmonton, AB T6G 2H4
Canada
E-mail: mcgee1@ualberta.ca

* The authors thank participants at annual meetings of the Canadian Economics Association, the Economic Science Association, the European Association of Labour Economists, the Society of Labor Economics, and the IZA World Labor Conference and seminar audiences at Simon Fraser, the University of Arkansas, and the University of Alabama for their suggestions. We thank Vera Brencic for kindly sharing her job ad data. Catherine Michaud-Leclerc and Hanh Tong provided outstanding assistance. We gratefully acknowledge the financial support from SSHRC Insight Development Grant #430-2013-00700. All remaining errors are our own.

1. Introduction

In the United States, 60-70% of employers in 2014 used online personality tests to screen applicants—spending \$500 million annually on such testing (Weber and Dwoskin [2014]). Firms screening job applicants with personality tests enjoy higher levels of productivity at the firm-level (Ichniowski et. al. [1997]) and among new hires (Autor and Scarborough [2008]). Employers presumably use personality tests to identify applicants with particular traits, but what employers learn from these tests is unclear as applicants have strong incentives to misrepresent themselves. When applying for a job, what applicant strongly agrees with the statement (from the online test of a major North American retailer), “When I encounter very difficult problems, I tend to move on to something else”? Surely applicants lie—probably a lot.

We investigate what can be learned from personality tests when individuals have incentives to misrepresent themselves and how employers use these tests in the labor market. We adopt Almlund et al.’s [2011] Roy model of personality testing in which test-takers allocate effort to manipulate responses to maximize the expected return to responses. The model highlights the role of incentives, information, and situations in the measurement of personality and how incentivized personality measures may be influenced by traits other than personality. For example, more intelligent applicants may be better at tilting responses in socially desirable ways or inferring what employers are looking for from cues. As such, incentivized personality tests may aggregate information about more than just personality.

To evaluate the information contained in incentivized personality tests, we conducted a within-subject, laboratory experiment with four experimental treatments. Subjects participated in two sessions. In the first “Baseline” session, subjects were paid \$7 to complete a questionnaire measuring, among other things, the Big 5 personality traits. No further incentives were provided.

About a week later, subjects participated in the second “Evaluation” session. Subjects were given a job ad and told they would complete personality and IQ tests after which a \$25 bonus would be awarded to the subject who best met the “hiring” criteria implicit in the job ad.

The four treatments differed in the job ads presented to subjects and the criteria used to award the bonus. In the *Extroversion* treatment, the job ad indicated that an extroverted worker was desired. We chose to prime a single trait, extroversion, because Hough and Oswald [2000] report that employers use personality tests specifically to identify extroverts. As a robustness exercise, the job ad in the *Introversion* treatment indicated that an introverted worker was desired. In the *No Priming* treatment, the ad provided little information about the type of worker desired. Finally, subjects in the *Audit* treatment were given the same ad as in the *Extroversion* treatment but were informed that they would be ineligible for the bonus if their answers in the Evaluation session differed from their Baseline answers by more than a given amount.

We present four findings from the experiment—in which subjects misrepresent themselves on all five personality traits in all treatments.¹ First, incentivized personality measures are only weakly to moderately correlated with non-incentivized personality measures in the *Extroversion*, *Introversion*, and *No Priming* treatments with correlations ranging between 0.2 and 0.6. In the *Extroversion* treatment, for instance, the Pearson correlation coefficient between the incentivized and non-incentivized extroversion measures is 0.38.

Second, we find little evidence that the non-Big 5 characteristics measured in our experiment—other than locus of control and perhaps risk aversion—are related to the

¹ By misrepresentation, we mean the difference between subjects’ trait scores in the Baseline and Evaluation sessions. The Baseline trait scores are themselves noisy measures of the underlying traits, but we interpret deviations from these scores in the primed or socially desirable direction as evidence of misrepresentation.

incentivized personality trait scores in all treatments as would be the case if these characteristics influenced either subjects' ability to misrepresent themselves or the costs of misrepresentation.

Third, we find evidence consistent with the hypothesis that more intelligent individuals are better at extracting information about desired personalities and adapting their responses accordingly. In the *Extroversion (Introversion)* treatment, a one standard deviation increase in IQ is associated with a 0.20 SD (0.18 SD) increase (decrease) in the incentivized extroversion score, but IQ is uncorrelated with incentivized extroversion scores in the *No Priming* treatment in which there is no information for subjects to extract. Likewise, we find little evidence that IQ is correlated with the personality traits that were not primed.

Finally, the incentivized and non-incentivized personality trait scores are more highly correlated in the *Audit* treatment than in any other treatment. The *Audit* treatment reduces the expected benefit to misrepresentation through the risk of detection, and subjects engage in considerably less misrepresentation as a result. We also observe a negative correlation between incentivized conscientiousness measures and IQ in the *Audit* treatment, which suggests intelligence may also influence incentivized personality measures if more intelligent individuals actively manage their misrepresentations in systematic ways so as to avoid detection.

A takeaway from the experiment is that responses to incentivized personality tests depend on the information available about the desired characteristics of a new employee, but how much such information does the average job ad contain? To what extent do employers signal their demand for workers with particular personality traits? Using a sample of 142,618 job ads posted to Monster.com over two weeks in 2006, we assess the personality traits demanded in job ads by counting the number of words associated with a trait in each ad. While the job ads are heterogeneous in terms of the extent to which they employ words associated with particular

traits, the median job ad contains 6 words associated with conscientiousness and 4 words associated with extroversion. Notably, our experimental job ads in the *Extroversion* and *Introversion* treatments would fall at the 94th and 71st percentiles of the distributions for the frequency of words associated with extroversion and introversion, respectively—suggesting that many job ads send signals concerning desired traits as strong as or stronger than those in our experiment. Overall, the evidence from the job ads indicates that many employers are looking for workers with particular personality traits.

The final issue we address is how employers actually use personality tests in the hiring process. On the one hand, our experiment indicates that employers could use personality tests as a poor substitute for intelligence tests. While easily administered and effective general mental ability tests exist (Schmidt and Hunter [1998]), employers in the United States, for example, may be wary of giving such tests when hiring if doing so exposes them to litigation risks in light of the *Griggs v. Duke Power Co.* (401 U.S. 424 [1971]) decision.² Firms may be happy to avail themselves of a single test that aggregates information on multiple dimensions. On the other hand, the evidence from the job ads suggests that employers are indeed looking for workers with particular traits, and the experimental results indicate that incentivized personality measures are correlated with non-incentivized measures.

Using the 2004 and 2011 rounds of the British Workplace Employment Relations Study, we present evidence that most employers use personality tests to complement performance or competency tests (that include intelligence tests) in hiring. Nearly 80 percent of establishments using personality tests when filling vacancies were also using performance or competency tests. Among the factors firms considered important in hiring (e.g., references, qualifications and

² The *Griggs v. Duke Power Co.* (401 U.S. 424) decision does not prohibit the use of IQ tests in hiring, but rather establishes that employers must show that such tests are necessary.

experience), only motivation and personality were correlated with the probability of administering a personality or attitude test when filling vacancies. Thus we infer that while incentivized personality tests may be informative about intelligence, firms appear to use these tests to learn about applicants' personalities, which is consistent with firms' expressed interest in workers with particular personality traits evident in the job ad sample.

Our findings underscore the identification challenges in measuring individual traits discussed in Almlund et al. [2011] and Borghans et al. [2011]—namely the need to standardize these measures to account for incentives and situations. To date, research on the role of incentives largely focuses on measures of cognitive ability in environments without strong incentives, which have been shown to be correlated with non-cognitive traits (e.g., Duckworth et al. [2011], Segal [2012], Heckman and Kautz [2012]). Our findings suggest that accounting for incentives and situations (here whether the threat that misrepresentations will be detected and punished is present) is of first-order importance when measuring personality, while accounting for the information available—though not without consequence—is less important.

As for the implications of our findings for empirical research on the role of personality traits in labor markets, researchers using personality measures from incentivized environments (e.g., Dal Bo et al. [2013]) can take heart in the fact that these tests are correlated with non-incentivized personality measures—especially when test-takers face some deterrent to misrepresentation as in our *Audit* treatment. At the same time, our findings imply the potential for an omitted variables bias when employing incentivized personality measures as a control. Specifically, when applicants have information about the employer's desired personality traits, personality measures may be correlated with IQ.

From the point of view of firms trying to best use personality testing in hiring, our findings are a mixed bag. Incentivized personality tests are likely very noisy measures of personality traits if their correlations with non-incentivized measures are anything to go by, and firms should appreciate the limits of using such tests to screen applicants. As we discuss in Section 5, using incentivized extroversion scores to trim the bottom half of the subject pool in order to eliminate less extroverted subjects as judged by their non-incentivized scores would do only marginally better than simply randomly eliminating subjects in the *Extroversion* treatment. At the same time, the *Audit* treatment shows that even a small threat that personality measures will be subject to verification as part of the hiring process may result in personality measures that are much more highly correlated with non-incentivized measures. Additionally, incentivized personality measures may help firms learn indirectly about applicants' IQ.

The remainder of the paper proceeds as follows. Section 2 reviews the psychology literature on personality testing in hiring and highlights the contributions of our study to this literature. Section 3 discusses how the Roy model of personality testing informs our research questions. Section 4 describes the experimental design, and Sections 5 and 6 present the results of the experiment and robustness checks. Section 7 discusses the information about desired personality traits in job ads on Monster.com. Section 8 presents evidence concerning how employers in the United Kingdom use personality tests when hiring. Section 8 concludes.

2. Personality testing and hiring

The development of instruments to study personality traits—the “enduring patterns of thoughts, feelings, and behaviors that reflect the tendency to respond in certain ways under certain circumstances” (Roberts [2009])—is nearly as old as psychology. The focus of this literature for the last 40 years has been the Five Factor Model or “Big 5” (Goldberg [1981]),

which organizes personality into five broad, fundamental traits: extroversion, conscientiousness, agreeableness, openness/intellect, and neuroticism (or emotional stability).³

The application of personality tests to employment decisions dates to at least the beginning of the 20th century (e.g., Freyd [1926]), and industrial/organizational psychologists have debated the usefulness of personality tests in hiring for nearly as long (e.g., Ghiselli and Barthol [1953]). Broadly speaking, psychologists' major concerns about personality testing as a selection tool involve faking and criterion validity.

That individuals can engage in “response distortion” or faking on personality tests has been extensively documented (e.g., Velicer and Weiner [1975]; Kroger and Wood [1993]; Holden and Hibbs [1995]). This literature has established several stylized facts such as that test-takers tilt their answers toward socially desirable responses (Ones and Viswesvaran [1998]; Paulhus [2002]; Donovan et al. [2003]), can fake both “good” and “bad” responses when directed to do so (Hough et al. [1990]; Furnham [1997]), and find it equally easy to fake all of the Big Five personality traits (Viswesvaran and Ones [1999]).

The faking research can be divided into two sets of studies. In the so-called “fake good” studies, subjects are asked to misrepresent themselves in (typically) favorable ways (e.g., Velicer and Weiner [1975], Furnham [1997], McFarland and Ryan [2000]). The second set of studies compares the personality test scores of incumbents to applicants for the same position (e.g.,

³ According to McCrae and John [1992], individuals high in extroversion tend to be more energetic, outgoing, ambitious, and assertive, while those low in extroversion are more reserved, quiet, and withdrawn. Individuals high in conscientiousness tend to be diligent, well-organized, neat, while those low in conscientiousness tend to be sloppier and more spontaneous. Individuals high in agreeableness tend to be more trusting, modest, and compliant, while those low in agreeableness tend to be more competitive and suspicious. Individuals high in openness/intellect have greater need for varied and novel experiences, greater aesthetic sensitivity, and greater curiosity, while those low in openness/intellect tend to be more pragmatic. Individuals high in neuroticism experience more depression, feelings of guilt, and anxiety, while those low in neuroticism are calm and even-tempered. The taxonomy used to describe the underlying traits varies across Big 5 instruments (e.g., “emotional stability” rather than neuroticism), but there is broad agreement about the five factor structure with some disagreement over the interpretation of which Big 5 factor subsumes more specific traits. See Almlund et al. [2009] for a comprehensive review of this literature.

Hough et al. [1990]; Becker and Colquitt [1992]; Rosse et al. [1998]; Birkeland et al. [2006]).

The “fake good” studies tend to find evidence of larger response distortions than the applicant-incumbent studies, but these studies typically explicitly direct subjects to fake without providing them with incentives to do so.⁴

Relative to these studies, our study makes a methodological contribution by using a within-subject experimental design among subjects with actual incentives (but not directives) to misrepresent themselves that allows us to assess the relationships between incentivized personality measures and non-incentivized personality measures and other characteristics. Unlike the between-subject applicant-incumbent studies or “fake good” studies without incentives, our study can estimate how non-personality traits influence faking and how incentives and faking affect the ranking of applicants by personality traits—an issue around which there is much debate (e.g., Morgeson et al. [2007 a,b], Ones et al. [2007]).

A study similar to ours is that of Tett et al. [2012], who use a within-subject, “fake good” design without incentives in which subjects were given job descriptions and asked to respond “in a way that will make you appear as an ideal job candidate.”⁵ They find that generalized intelligence explains some of the observed faking (as evidenced by statistically significant changes in R^2), but their study was not designed to establish the conditions under which intelligence affects incentivized personality measures. By contrast, our study establishes that without verification warnings, IQ influences incentivized personality traits only when there is information about desired personalities to be extracted and only for those traits being demanded.

⁴ Exceptions include Dwight and Donovan [2003] and Vasilopoulos et al. [2005], discussed below, who used between-subject, “fake good” designs in which the top-performing subjects were paid bonuses as in our experiment.

⁵ Tett et al. did not attempt to explicitly prime particular traits in their job descriptions. Other “fake good” studies use job titles to direct subjects to fake as they think a certain type of worker (librarian, banker, salesman) would (e.g., Velicer and Weiner [1975], Furnham [1990]), but these studies rely on stereotypes about professions rather than providing information about desired characteristics.

Moreover, by providing incentives *without explicitly directing subjects to fake*, the response distortions in our experiment may be more similar to those of actual job applicants than in “fake good” studies. Finally, while Tett et al. speculate that traits such as Machiavellianism and dishonesty may influence faking, our study measures these traits and tests these hypotheses.⁶

The recognition that faking poses serious problems in selection contexts led to a large literature devoted to the detection of faking on personality tests and potential corrections, but whether faking can be detected and trait measures corrected remains controversial (e.g., Goffin and Christiansen [2003], Rothstein and Goffin [2006]). A related literature considers the efficacy of “verification warnings” where test-takers are warned that their responses are subject to verification in one way or another. Using mostly between-subject designs, these studies suggest that verification warnings reduce faking. Dwight and Donovan [2003], for instance, find that personality measures in a between-subject study are least inflated when subjects are warned that faking is detectable and informed of the consequences of detection—both characteristics of our *Audit* treatment. Relative to the between-subject studies, our within-subject design allows us to address the extent to which the threat of detection in the *Audit* treatment can offset the competing effects of incentives for misrepresentation in terms of whether the incentivized extroversion scores rank subjects from most to least extroverted more similarly to the non-incentivized measures than in the *Extroversion* treatment.

Within the “verification warning” literature, perhaps most relevant for our study is Vasilopoulos et al. [2005], who find positive correlations between cognitive ability and conscientiousness measures when applicants for law enforcement positions are warned that their

⁶ Kleinmann et al. [2011] also speculate that the link between incentivized personality test scores and productivity may result from applicants differing in their ability to identify the criteria used to evaluate their responses if this ability is related to productivity, but they provide no evidence to this effect.

responses may be subject to verification but no correlations in the absence of these warnings. Similarly they find that SAT scores are positively correlated with conscientiousness scores among undergraduates imagining that they are applying for admission to a desirable university when these students are warned that their responses would be subject to verification from teachers, friends, and past employers but not among subjects without this warning. In both studies, Vasilopoulos et al. find that cognitive ability is related to response times under the warning condition, suggesting that more intelligent individuals exert more effort responding under these conditions.

In our *Audit* treatment, by contrast, we find that IQ is negatively correlated with incentivized conscientiousness scores. We speculate that the difference between our studies may lie in the nature of the verification threat. Individuals in the Vasilopoulos et al. studies were given verification warnings, whereas subjects in our study were informed of the procedure that would be used to audit their responses. As a consequence, more intelligent subjects in our study may have devoted more effort deciding where to reign in their faking, whereas more intelligent individuals in the Vasilopoulos et al. studies may have devoted more effort identifying less easily verified items for faking as Donovan et al. [2003] suggest individuals do. Overall though, we view our findings and those of Vasilopoulos et al. as indicative of a role for intelligence in influencing incentivized personality measures when verification warnings are present.

The second major concern among psychologists regarding the use of personality tests in hiring centers around their criterion validity, the degree to which measured personality traits are correlated with job performance. While there is a consensus that Big 5 traits are related to job

performance, the estimated correlations tend to be modest and the subject of debate (e.g., Barrick and Mount [1991]; Christiansen et al. [1994]; Hurtz and Donovan [2000]).⁷

Our study offers potential explanations for why criterion validity estimates vary by demonstrating the importance of the information about desired personalities and the presence of verification warnings to the correlations between incentivized personality measures and non-incentivized personality measures and IQ. Variation across jobs in the information about desired traits provided may result in differences in criterion validity estimates if incentivized personality measures are correlated with productivity because of their correlation with IQ for some jobs but not others. Likewise, variation across firms in the use (and credibility) of verification warnings may result in differences in criterion validity estimates if verification warnings result in considerably more precise measures of personality traits.

3. Modeling Behavior on Incentivized Personality Tests

In Almlund et al.'s [2011] Roy model of personality testing, personality tests consist of items P_j ($j = 1, \dots, J$) with payoffs R_j for each item. Performance (i.e., a response) is assumed to be a function of individual characteristics (θ) and the effort allocated to the item, e_j :

$$P_j = \phi_j(\theta, e_j)$$

Individuals are endowed with total mental effort \bar{e} that they allocate across items. The cost of expending effort on item j is given by $C_j(\theta, e_j)$. Test-takers allocate effort to maximize the expected returns to performance given the available information \mathbb{I} subject to the effort constraint:

$$\max_{\{e_j\}_{j=1}^J} E\{\sum_{j=1}^J [R_j \phi_j(\theta, e_j) - C_j(\theta, e_j)] \mid \mathbb{I}(\theta)\} \text{ subject to } \sum_{j=1}^J e_j = \bar{e}$$

⁷ The correlations between Big 5 traits and job performance measures range from 0 to approximately 0.15 (see Table 2 in Morgeson et al. [2007b] for a useful summary of several meta-analyses).

The notion that some traits may be related to “faking” is captured by either the cost function or the production function ϕ_j . Specifically, the costs of manipulating performance through effort might be decreasing in some traits. That is, $\frac{\partial^2 c_j(\theta, e_j)}{\partial e_j \partial \theta} < 0$. For instance, intelligent individuals may find exerting mental effort less onerous. Alternatively, some individuals may dislike misrepresenting themselves less than others. Note that the second example highlights the fact that θ may include preference parameters as well as traits like personality and intelligence. As Almlund et al. note, allowing the psychic costs of effort to be increasing or decreasing in θ is empirically indistinguishable from the assumption that traits θ and effort e_j are complements or substitutes in the performance of task j (i.e., $\frac{\partial^2 \phi_j(\theta, e_j)}{\partial e_j \partial \theta} \geq 0$).

Individuals maximize the expected returns given their information set $\mathbb{I}(\theta)$, where the uncertainty surrounds the returns to tasks R_j . The information set $\mathbb{I}(\theta)$ may depend on individual characteristics θ as well as the situation if, for instance, more intelligent or socially astute individuals are better at extracting cues from situations than others.

Almlund et al. observe that in order to identify personality traits from performance on test items P_j one needs to standardize for effort levels, incentives, situations, information sets, etc. For simplicity, assume that truthful reporting on test item j implies that $e_j = 0$.⁸ Thus “faking” involves the manipulation of performance on the test item through effort ($e_j > 0$). The model gives rise to the following research questions concerning the information contained in incentivized personality tests such as those used in hiring:

Research question #1: How important are incentives when measuring personality?

⁸ If test-takers have no incentive to misrepresent themselves, one might assume that $R_j = 0$ for all j , and subjects maximize their returns to performance by minimizing effort costs $C_j(\theta, e_j)$ by setting $e_j = 0$ for all j .

Our within-subject experiment obtains the same personality measures from the same subjects in two environments: one without incentives (i.e., $R_j = 0$) and one with powerful incentives for misrepresentation (i.e., $R_j \neq 0$). Assessing the correlation between personality measures with and without incentives for misrepresentation sheds light on the empirical importance of standardizing for incentives and effort when measuring personality.

Research question #2: Are incentivized personality measures influenced by traits other than personality? Is the ability to fake or the cost of faking related to other characteristics?

In the model, the ability to fake or the costs of faking may depend on traits other than personality. As a consequence, incentivized personality measures may be correlated with measures of traits with which they are uncorrelated in the absence of faking (i.e., when $e_j = 0$). A subtler point is that if a correlation arises between a personality measure and another measured characteristic because the latter makes exerting effort faking more productive or less costly, then a correlation should be observed in all situations that induce faking.⁹

Research question #3: How does the information available to test-takers influence the relationship between measured personality and other traits?

Some traits may influence the information individuals extract from their environments and thus expectations about the returns to task performance R_j . As such, whether correlations exist between a measured characteristic and incentivized personality measures may depend on the information available.¹⁰

⁹ Consider two uncorrelated traits, θ_1 and θ_2 , and a test measuring θ_1 consisting of a single item P_1 . Let performance on the item be given by $P_1 = \alpha_1\theta_1 + \alpha_2\theta_2 + \alpha_3e_1 + \alpha_4e_1\theta_2$. If $\alpha_2 = 0$, then P_1 will be uncorrelated with θ_2 in the absence of faking (i.e., when $e_1 = 0$). When a complementarity between effort e_1 and trait θ_2 exists (i.e., $\alpha_4 \neq 0$), the personality measure P_1 will be correlated with θ_2 even if $\alpha_2 = 0$ whenever faking occurs (i.e., $e_1 > 0$).

¹⁰ Suppose $P_1 = \alpha_1\theta_1 + \alpha_3e_1$. That is, trait θ_2 affects the personality measure P_1 neither directly (because $\alpha_2 = 0$) nor through faking ability (because $\alpha_4 = 0$). If the return to task performance R_1 is known or all subjects have the same expectations concerning this return because no information is available, then measured personality $P_1 = \alpha_1\theta_1 + \alpha_3e_1^*(R_1)$ will be uncorrelated with θ_2 where $e_1^*(R_1)$ is the optimal effort on task 1 given the known or

Research question #4: If faking on personality tests may be detected and punished, how much more correlated will personality measures with incentives for misrepresentation be with personality measures without such incentives than in the absence of such scrutiny?

In practice, job applicants may be aware—and possibly warned—that firms attempt to detect faking and reign in their response distortions. Alternatively, applicants may be wary of faking on a personality test only to be exposed as a fraud during a subsequent interview when their personality is also assessed.¹¹ The threat of exposure could lead to incentivized personality measures more similar to those obtained in the absence of faking.¹²

4. Experimental Design & Data

The experiment employed a within-subject design over two sessions, the Baseline and Evaluation sessions. We briefly describe the sessions before detailing the treatments.¹³

4.1 Baseline Session

In the Baseline session, subjects answered a questionnaire consisting of a 100-item Big 5 assessment (DeYoung [2007]), the Balanced Inventory of Desirable Responding (Paulhus [1984]), a self-monitoring scale (Snyder [1974]), a Machiavellianism scale (Jackson [1994]), an optimism-pessimism scale (Scheier et al. [1994]), questions about the acceptability of lying and

common expected return to item 1, R_1 . If instead R_1 is unknown and θ_2 affects individuals' inference concerning the return to item 1, then $P_1 = \alpha_1 \theta_1 + \alpha_3 e_1^*(E[R_1 | \theta_2])$ and P_1 may be correlated with θ_2 even if θ_2 has no effect on the ability to fake or costs of faking.

¹¹ Appeals for honesty are also widely used in the selection process. For example, Dal Bo et al. [2013] asked applicants for Mexican civil service positions to sign an honor code and indicated that prior employers would be called prior to administering a personality test.

¹² Imagine a test with a single item and known returns to performance. Suppose the event of detection is an increasing function of effort $D(e_j)$ and the probability of detection is given by $\Pr(D(e_j))$. If detected, subjects earn zero return. Test-takers maximize the expression

$$[1 - \Pr(D(e_j))]R_j \phi_j(\theta, e_j) - C_j(\theta, e_j)$$

Effectively, the threat of detection reduces the expected return to faking.

¹³ The experiment was programmed and conducted using z-Tree (Fischbacher [2007]).

demographic questions.¹⁴ Subjects were informed at the outset that they would only receive a \$7 payment for participation. As such, their answers were not explicitly incentivized.

The Baseline session lasted 15-20 minutes for most subjects. The instructions informed subjects that their participation made them eligible for a future experiment. After paying each subject, our research assistant offered to sign the subject up for the Evaluation session at the same time the following week. In most cases, subjects agreed, but subjects were not required to participate in the additional session and not all did. As a result, most subjects participated in the Baseline and Evaluation sessions one week apart, but nothing prevented subjects from registering for the Evaluation session more than one week after the Baseline.

4.2 Evaluation Session

In the Evaluation session, subjects were placed in groups of four subjects.¹⁵ Subjects did not interact with other members of their group and were not aware which subjects were in their group. The instructions informed subjects that they would receive a job description and then complete personality and intelligence tests and that a \$25 bonus would be awarded to one member of each group on the basis of the tests. The job description and what subjects learned about how the tests would be used to award the bonus depended on the treatments. See Appendix A for copies of the instructions and job descriptions for each treatment.

After the instructions, subjects had 20 minutes to complete Raven's Progressive Matrices (RPM, Raven [1998]), an intelligence test. They then had as much time as they needed to answer

¹⁴ DeYoung's [2007] 100-item Big 5 assessment asks subjects to indicate how well a statement describes them using a 5-point Likert scale ranging from "Strongly disagree" to "Strongly agree." For instance, "I start conversations" is an item on the extroversion scale, while "I get stressed out easily" is an item on the neuroticism scale. For positively scored items, "Strongly disagree" is given a value of 1 and "Strongly agree" a value of 5; the values are reversed for negatively scored items. Each trait score is the sum of the values for items associated with that trait. Each trait is associated with 20 items, resulting in trait scores between 0 and 100.

¹⁵ When the number of subjects was not a multiple of four, some groups had three or five subjects.

the same 100-item Big 5 battery completed in the Baseline session (albeit with the items in a different order). Subjects then learned whether they had received the bonus before completing the incentivized Holt and Laury [2002] risk preference measure. Finally, subjects completed a short questionnaire measuring self-deception, impression management and locus of control and answered the same demographic questions as in the Baseline session. Subjects were also asked whether they were employed, how many jobs they had worked, how many years of work experience they had, how many times they had taken a personality test as part of a job application, how confident they were that they knew what employers were looking for when looking for a job, and how the job description had affected their behavior in the experiment. In addition to a \$7 participation fee and (if applicable) the \$25 bonus, subjects were paid \$0.20 for each correct answer on the RPM and their earnings from one randomly selected realization of a paired lottery of the Holt-Laury instrument.

4.3 Treatments

Extroversion treatment: The *Extroversion* instructions informed subjects that “All members of a group will receive the same job description. We will then administer an intelligence test and a personality test to determine who to “hire” for the job. We will weigh these two tests according to some criteria and one subject within each group who best meets these criteria will receive a bonus of \$25; the remaining subjects in each group will not receive any bonus.” The job description indicated that a staffing firm was looking to hire an extrovert for a “client services representative” position. To indicate that an extrovert was desired, the job description incorporated words and phrases associated with extroversion such as “outgoing,” “able to take initiative, be assertive,” and “proficient at building and maintaining relationships” while attempting to avoid indicating that the firm was seeking workers with other traits.

Introversion treatment: The *Introversion* instructions were the same as in the *Extroversion* treatment, but the job description indicated that the staffing firm was looking for an introverted client services representative. Specifically, the job description indicated that a “low-key” firm was looking for a “contemplative, reserved, independent” individual who “enjoys tackling projects solo” and who gets “the job done without making a splash or interrupting clients’ normal business.”

No Priming treatment: The instructions in the *No Priming* treatment were again the same as those in the *Extroversion* treatment. The job description in the *No Priming* treatment was also for a “Client Services Representative” at a staffing firm, but the job description otherwise contained very little detail. The description indicated requirements such as “proficient in Microsoft Office” and “has financial acumen”—requirements that were also in the *Extroversion* and *Introversion* job descriptions. The position title (“Client Services Representative”) and the associated tasks may themselves have “primed” subjects based on prior beliefs, but we attempted to provide as little additional priming as possible. Where the *Extroversion* job description was 272 words long, the *No Priming* job description was just 99 words.

Audit treatment: Subjects in the *Audit* treatment received the same job description used in the *Extroversion* treatment. After explaining that the personality and intelligence tests would be used to award the bonus (as in the other treatments), the instructions informed subjects that there was a 50 percent chance that their personality test answers would be compared to their answers from the Baseline session. In these audits, answers from eight questions were compared across sessions; subjects did not know which questions would be spot-checked. If the Evaluation session answers differed by more than a certain amount in the aggregate from the Baseline responses, subjects would fail the audit and be ineligible for the bonus. See the instructions in

Appendix A for the description of the audit criteria provided to subjects. If the subject who won the bonus was audited and failed the audit, no bonus was awarded within the group.

Test-Retest treatment: Subjects in the *Test-Retest* treatment completed two Baseline sessions one week apart. The test-retest reliability of a psychometric instrument—the extent to which the instrument produces similar measures when taken at different points in time—depends on how far apart in time subjects respond to the instrument. Because most subjects completed the Baseline and Evaluation sessions seven days apart, the *Test-Retest* treatment provided us with a benchmark against which to compare our findings from the other treatments.

4.4 Subject characteristics and hypotheses

We now describe the measures of subject characteristics and the hypotheses that motivated our collection of these measures. The survey instruments are provided in Appendix A.

IQ: We hypothesized that more intelligent individuals might be better at adapting their responses toward socially desirable responses. Alternatively, more intelligent individuals might be better at recognizing cues concerning desired personality traits and responding accordingly. To measure fluid intelligence, we used the RPM test.¹⁶ Items on this test consist of a picture of a pattern with a section missing. Subjects choose among possible patterns to complete the missing section. The IQ score is the sum the number of correct answers.

Machiavellianism/Self-monitoring/Self-deception/Impression management: We hypothesized that some individuals may differ in their willingness to misrepresent themselves—preferences incorporated in the effort cost function in Section 3—and that some characteristics—specifically Machiavellianism, self-monitoring, self-deception, and impression management—

¹⁶ Intelligence can be thought of as having a fluid component and a crystallized component. An individual's fluid intelligence is their ability to solve a novel problem when they have little existing knowledge relevant to the problem. Crystallized intelligence, on the other hand, is the ability to recall and apply existing knowledge gained through experience or education to problem solving (Cattell [1963]).

may be correlated with this willingness to engage in misrepresentation. The Machiavellianism measure assesses the degree to which a respondent agrees that he/she can manipulate or deceive others through agreement or disagreement with statements like “I can talk others into doing things.” The self-monitoring instrument measures the extent to which respondents pay attention to and control their expressive behavior and self-presentation in accordance with what is socially acceptable through agreement with statements such as “I put on a show to impress people.” Self-deception involves the degree to which respondents believe positive statements about themselves that are not obviously true or false such as “I always know why I like things.” Impression management measures the conscious effort by a respondent to present themselves in a positive light through agreement with statements such as “I have never dropped litter on the street.”

Acceptability of lying: We hypothesized that preferences for honesty would be negatively correlated with misrepresentations. To measure preferences for honesty in different domains, subjects in the Baseline session were asked to rate the acceptability of lies in 16 scenarios: four scenarios each describing individuals lying for personal gain, lying to avoid conflict, lying to gain social acceptance, or lying to be kind (McLeod and Genereux [2008]). Responses concerning each scenario used a 9-point Likert scale ranging from “Extremely unacceptable” (assigned a score of 1) to “Extremely acceptable” (assigned a score of 9). We measure willingness to lie by motivation category by summing the four responses within each category.

Optimism-pessimism: We hypothesized that optimistic subjects might feel less need to misrepresent themselves if their optimism extended to their beliefs about the probability of winning the bonus. The optimism scale measures how optimistic subjects are through agreement with statements such as “In uncertain times, I usually expect the best.”

Risk aversion: We hypothesized that risk averse individuals would be less willing to engage in misrepresentations in the *Audit* treatment when misrepresentations risked being detected and rendering a subject ineligible for the bonus.¹⁷ To measure risk aversion, we employ the Holt-Loury risk preference measure. Subjects choose between a “safe” lottery and a risky lottery with the potential for higher payouts. The riskiness of the latter option increases over ten paired lotteries. Risk aversion is measured by the number of “safe” lotteries chosen.

Locus of control (internality): We hypothesized that individuals with an internal locus of control who believe that outcomes are primarily a function of their own efforts would exert more effort misrepresenting themselves. In a search experiment, McGee and McGee [2016] show that search effort is positively correlated with internality when the return to effort is uncertain (and thus beliefs are informed by locus of control) but uncorrelated with internality when the return to effort is known. In the current study, subjects were not explicitly asked to misrepresent themselves, nor could subjects have been certain how these misrepresentations would be rewarded. Thus we expected internal individuals—who might believe the return to faking effort to be higher—to misrepresent themselves more. Locus of control beliefs (Rotter [1966]) are measured through agreement with statements such as “What happens to me is my own doing.”

4.5 Summary statistics and balance checks

A total of 642 subjects participated in both a Baseline and Evaluation session: 167 subjects in the *Extroversion* treatment, 160 in the *Introversion* treatment, 147 in the *No Priming* treatment, and 168 in the *Audit* treatment. Another 45 subjects participated in the *Test-Retest*

¹⁷ In section 3, we assumed test-takers maximize a linear payoff function, effectively removing any role for risk aversion. One could, of course, employ a non-linear utility function to introduce a role for risk preferences.

treatment. All of the subjects were undergraduates at Simon Fraser University except for those in the *Test-Retest* treatment, which was run at the University of Arkansas.

Table 1 presents means and standard deviations for the non-Big 5 measures for each treatment (excluding the *Test-Retest* treatment). The last column reports p-values for Kruskal-Wallis tests for differences in characteristics across the treatments to assess whether the randomization of subjects to treatments was successful. The only significant differences across treatments are in the age of participants ($p = 0.007$) and IQ ($p = 0.057$).¹⁸

Table 2 reports the means by treatment for the Big 5 trait scores in the Baseline and Evaluation sessions; the last column again reports the p-values for Kruskal-Wallis tests for differences across the treatments. In the Baseline sessions, none of the differences across treatments in Big 5 traits are statistically significant at conventional levels. Overall the random assignment of subjects to treatments appears to have been largely successful.¹⁹

5. Experiment results

Subjects in the experiment unambiguously respond to pecuniary incentives to misrepresent themselves—which is noteworthy insofar as we did not explicitly ask subjects to misrepresent themselves as in the “fake good” studies. The p-values for the nonparametric Wilcoxon signed-rank tests of the equality of the mean trait scores in the Baseline and Evaluation sessions in Table 2 are less than 0.001 for each trait in every treatment with the exception of openness/intellect

¹⁸ We also find statistically significant differences in optimism and impression management in the Evaluation session. Because these measures may have been influenced by the treatments and were also measured in the Baseline session, we use the Baseline measures in the analysis.

¹⁹ Subjects were not obliged to participate in the Evaluation session, and 162 subjects completed only the Baseline session. To evaluate whether our sample of returnees is systematically different from non-returnees, we compare the characteristics of the two groups in Appendix B Table 1. Subjects who did not return for an incentivized session were slightly more extroverted than those who returned (Wilcoxon rank-sum $p = 0.098$) and engaged in slightly less self-deception ($p = 0.091$), but otherwise we find no significant differences between the two groups.

(0.337) and agreeableness (0.192) in the *Audit* treatment.²⁰ In most cases, the Baseline and Evaluation session means differ by a full standard deviation relative to the Baseline distribution.

Turning to the distributions themselves, Figure 1 displays the kernel densities of the extroversion scores from the Baseline and Evaluation sessions in all four treatments; Appendix B Figure 1 displays the same kernel densities for the other personality traits. Kolmogorov-Smirnov tests reject the null hypothesis of equality of the incentivized and non-incentivized distributions for every trait in the *Extroversion*, *Introversion*, and *No Priming* treatments at the 5% significance level. In the *Audit* treatment, we fail to reject the null at the 5% level for every trait except neuroticism—although at the 10% significance level we would reject the null of equality of the distributions for extroversion, conscientiousness, and neuroticism. We now consider our findings as they relate to the research questions detailed in Section 3.

Research question #1: How important are incentives when measuring personality?

If subjects misrepresent themselves by a common amount or proportion, the shifts in the distributions in Figure 1 would be order-preserving insofar as the ranking of subjects from most to least extroverted would be unchanged. Figure 2 displays the scatterplots by treatment of the Baseline extroversion scores against the Evaluation extroversion scores.²¹ The scatterplots reveal considerable heterogeneity in the extent to which subjects misrepresent themselves. Subjects with low Baseline extroversion scores have higher Evaluation extroversion scores than subjects with higher Baseline extroversion scores in all treatments—indicating that the treatment effects are not order-preserving.

²⁰ We reach the same conclusions using Benjamini and Hochberg [1995] q-values as implemented in Anderson [2008] to account for the testing of multiple hypotheses. These q-values are available from the authors.

²¹ Scatterplots for the other personality traits are provided in Appendix B Figure 2.

Panel A of Table 3 reports the pairwise Pearson correlation coefficients between the Baseline and Evaluation scores for each of the Big 5 personality traits by treatment. In the *Extroversion*, *Introversion*, and *No Priming* treatments, these correlations are weak to modest—particularly relative to the *Test-Retest* treatment. In the *Extroversion* treatment, for example, the correlation between the Baseline and Evaluation extroversion scores is just 0.38 as compared to 0.94 in the *Test-Retest* treatment.

Panels B and C report Spearman's rank correlation and Kendall's tau-b coefficients, respectively. Both statistics measure the extent to which the Baseline and Evaluation scores are related in a monotonic (i.e., order-preserving) manner. For almost every trait in all treatments, the Spearman rank correlation and Kendall's tau-b are lower than the Pearson correlation coefficients. In the *Extroversion* treatment, the Spearman correlation between the Baseline and Evaluation extroversion scores is 0.31 as compared to 0.93 in the *Test-Retest* treatment.

The correlations have two implications. First, empirical researchers using incentivized measures of personality traits such as those generated in the hiring process should recognize that these are likely very noisy measures of the underlying trait if their correlations with non-incentivized measures are anything to go by. Standardizing personality measures for incentives appears to be very important. Second, firms using personality tests to trim the applicant pool (as Autor and Scarborough [2008] and Hoffman et al. [2018] suggest that firms do) should recognize that the ranking of applicants by traits generated by such tests is likely a weak approximation of the true ranking. To fix ideas, suppose firms used personality tests to eliminate applicants in the bottom half of the extroversion distribution, and assume for the moment that non-incentivized scores are error-free measures of the underlying traits. In the *Extroversion* treatment, 51 of the 84 subjects in the bottom half of the non-incentivized extroversion distribution would be eliminated

if employers trimmed this applicant pool based on incentivized extroversion scores. By contrast, eliminating the bottom half of subjects when ranking them randomly would eliminate on average 42 of the 84 subjects in the bottom half of the non-incentivized extroversion distribution.

Research question #2: Are incentivized personality measures influenced by traits other than personality? Is the ability to fake or the cost of faking related to other characteristics?

The scatterplots in Figure 2 suggest heterogeneity in the extent to which subjects misrepresent themselves. Some of this heterogeneity may result from differences in characteristics that influence subjects' ability to fake or the costs of faking. To evaluate whether other traits influence the incentivized extroversion scores, we regress for each treatment the standardized, incentivized extroversion score on the standardized, non-incentivized extroversion score and standardized measures of the characteristics discussed in Section 4.4. The non-incentivized and incentivized extroversion scores are standardized using the Baseline and Evaluation session extroversion scores, respectively, of subjects in the same treatment. The other trait measures are standardized using the full sample of scores for subjects in all treatments.

Columns (1) to (4) of Table 4 report the OLS coefficient estimates from these regressions for the *Extroversion*, *Introversion*, *No Priming*, and *Audit* treatments, respectively. The heteroskedasticity-robust standard errors reported in parentheses are re-scaled by the square of the observation's estimated variance as recommended by Davidson and MacKinnon [1993] given our sample sizes.²² Similar to the Pearson correlation coefficients in Table 3, the

²² Subjects with high Baseline extroversion scores have little scope to increase their incentivized extroversion scores, while subjects with low Baseline extroversion scores have considerable scope for misrepresentation, resulting in significant heteroskedasticity.

coefficients of the non-incentivized Extroversion score range from 0.413 in the *Extroversion* treatment to 0.632 in the *Audit* treatment.²³

In Section 3, we noted that if subject characteristics influence the ability to fake or the costs of faking for a given personality trait, then these characteristics should be correlated with incentivized trait scores in any treatment in which subjects exert effort to misrepresent themselves. With this in mind, Column (5) reports the p-values from Wald tests of the joint significance of the coefficients of the characteristic in a given row across the *Extroversion*, *Introversion* and *No Priming* treatments.²⁴ We reject the null that the coefficients in the *Extroversion*, *Introversion* and *No Priming* treatments are jointly equal to zero for only non-incentivized extroversion and IQ. A one standard deviation increase in IQ is associated a 0.196 standard deviation increase in the incentivized extroversion score in the *Extroversion* treatment and a 0.178 standard deviation reduction (i.e., an increase in introversion) in the *Introversion* treatment. Notably, IQ is not correlated with incentivized extroversion in the *No Priming* treatment as it should be if more intelligent individuals are simply better at faking extroversion.

More surprisingly, risk aversion is nearly jointly significant across treatments with a one standard deviation increase in risk aversion associated with a 0.153 standard deviation increase in incentivized extroversion scores in the *Extroversion* treatment and smaller and less precisely increases in the *Introversion* and *No Priming* treatments. We hypothesized that risk averse subjects may lie less in the *Audit* treatment because of the risk of being audited, but perhaps risk

²³ Given the manner in which we standardize the incentivized and non-incentivized extroversion scores, the coefficients of non-incentivized extroversion in Table 4 would equal the Pearson correlation coefficients in Table 3 in the absence of other controls.

²⁴ We exclude the *Audit* treatment from the test of joint significance because this treatment introduced a deterrent to the faking that the characteristics were hypothesized to influence. In fact, the inferences from the tests of the joint significance of the treatment-specific coefficients are unaffected for each trait when the *Audit* treatment coefficients are included in the null hypothesis. We implement the tests by estimating fully interacted models using data from all treatments with the controls interacted with treatment indicators.

averse subjects attempt to minimize the risk of not winning the bonus by exerting more effort to fake in socially desirable directions.

The remaining subject characteristics (locus of control, self-deception, impression management, willingness to lie for personal gain, Machiavellianism, and optimism-pessimism) are not significantly correlated with incentivized extroversion scores in any treatment, nor are any of the controls jointly significant across treatments. None of these measured characteristics appear to influence subjects' misrepresentations where extroversion is concerned.²⁵

Our focus on incentivized extroversion scores in Table 4 is motivated by Research Question 3 discussed below, but in fact the hypothesis that subject characteristics may influence the ability to fake or the costs of faking applies equally to all of the Big 5 traits. Table 5 reports abridged coefficient estimates for similar regressions by treatment using incentivized conscientiousness, agreeableness, openness, and neuroticism as the dependent variables; the full coefficient estimates are given in Appendix B Table 2. In contrast to the incentivized extroversion regressions, the coefficient of IQ is jointly significant across the *Extroversion*, *Introversion*, and *No Priming* treatments for a single trait: openness. Self-deception, willingness to lie for personal gain, Machiavellianism and optimism are again each statistically insignificant in every regression, and the coefficients of each measure in the *Extroversion*, *Introversion*, and *No Priming* treatments are jointly insignificant for the four additional personality traits. Risk aversion is positively correlated with incentivized agreeableness in the *No Priming* treatment while impression management is negatively correlated with incentivized agreeableness in the

²⁵ We include among the controls in Table 4 the measure of willingness to lie for personal gain as this seemed to correspond most closely to preferences for honesty in this context. We measured willingness to lie in four domains, but none of these measures were correlated with the incentivized personality scores. Likewise, we used principal component analysis to construct a common factor score from the willingness to lie measures as well as for the self-deception, impression management, and Machiavellianism measures, but these index variables were also uncorrelated with the incentivized personality scores.

Introversion treatment, but otherwise risk aversion and impression management are uncorrelated with the incentivized personality trait measures.

Only locus of control appears to be broadly correlated with incentivized measures of personality traits. Specifically, internality is significantly correlated with the incentivized conscientiousness, openness and neuroticism measures in the *Introversion* treatment and with incentivized conscientiousness in the *No Priming* treatment while being nearly significantly correlated with incentivized trait measures in several other treatments. The locus of control coefficients are jointly significant across treatments for the incentivized conscientiousness and neuroticism measures and nearly so for openness and agreeableness. Thus we find some evidence in support of the hypothesis that more internal individuals supply more effort in misrepresenting themselves than less internal individuals. Overall we find that only IQ, risk aversion and locus of control may be related to the ability to fake or the cost of faking.

Research question #3: How does the information available to test-takers influence the relationship between measured personality and other traits?

We first consider whether the information available to test-takers influences the correlations between incentivized and non-incentivized personality measures. Specifically, we test whether the Pearson correlation coefficients for the *Extroversion* and *Introversion* treatments equal those in *No Priming* treatment. Table 3 reports both unadjusted p-values for these two-sided tests and Holm-Sidak corrected p-values accounting for the 15 hypotheses tested in the table. The correlations between the non-incentivized and incentivized personality measures in the *Extroversion* treatment are lower than those in the *No Priming* treatment, and the unadjusted p-values would provide some evidence using one-sided tests that this difference is statistically significant at the 10% level for extroversion, openness and neuroticism. The multiple hypothesis

corrected p-values, however, make it clear that we fail to reject the null that the correlations are equal in the *Extroversion* and *No Priming* treatment and similarly for the *Introversion* and *No Priming* treatment, where the correlations are basically identical except for conscientiousness. Thus we find little evidence that the provision of information about desired personality traits influences the correlations between incentivized and non-incentivized personality measures.

This does not mean that the provision of information about desired personalities has no effect on incentivized personality measures. If subjects with particular traits are better at extracting information from cues concerning desired personality, then these traits may be correlated with incentivized personality measures when there is information to be extracted but uncorrelated with incentivized personality measures in the absence of such information.

For every characteristic other than IQ, we fail to reject the null hypothesis that the coefficients in the *Extroversion* and *No Priming* treatments in Columns (1) and (3) in Table 4 are equal; we likewise fail to reject the equality of the coefficients in the *Introversion* and *No Priming* treatments in Columns (2) and (3) for every characteristic except IQ. For IQ, we reject these nulls at the 1% significance level. Indeed, we fail to reject the null that the IQ coefficient in the *Extroversion* treatment has the same magnitude as that in the *Introversion* treatment but with opposite sign. We view the fact that the correlation between IQ and incentivized extroversion reverses sign in the *Extroversion* and *Introversion* treatments as strong evidence that this correlation is influenced by the treatment—the availability of information. By contrast, IQ is uncorrelated with the incentivized extroversion score when there is no information about the ideal personality type to be extracted in the *No Priming* treatment (Column (3)).

Regarding the relationship between IQ and the personality traits that were not primed, we fail to reject the null hypothesis that the IQ coefficients in Table 5 are equal to one another in the

Extroversion, *Introversion*, and *No Priming* treatments when all of the remaining incentivized Big 5 traits except incentivized openness serve as the dependent variable; for incentivized openness we reject this null at the 5% level (Panel C).²⁶ Overall, the lack of correlations between IQ and the non-primed trait measures is consistent with the hypothesis that IQ influences incentivized personality measures through the information acquisition or usage channel.²⁷

Research question #4: If faking on personality tests may be detected and punished, how much more correlated will personality measures with incentives for misrepresentation be with personality measures without such incentives than in the absence of such scrutiny?

The *Audit* treatment introduces the possibility that misrepresentations may be detected and punished when subjects are given the same job description as in the *Extroversion* treatment. In short, the average misrepresentations in the *Audit* treatment are much reduced as evidenced by the means in Table 2 and the kernel densities in Figure 1. As a consequence, the correlations between the incentivized and non-incentivized personality measures are significantly higher in the *Audit* treatment than in the *Extroversion* treatment. The Pearson correlations in Table 3 range between 0.59 to 0.78 in the *Audit* treatment compared to between 0.3 and 0.55 in the *Extroversion* treatment. For three of the Big 5 traits, we reject the null that the correlations in

²⁶ Likewise we fail to reject the equality of the coefficients in the *Extroversion*, *Introversion* and *No Priming* treatments for all of the remaining subject characteristics—including risk aversion and locus of control—when incentivized conscientiousness, agreeableness, openness/intellect, and neuroticism serve as the dependent variables. This suggests that the relationships between risk aversion and locus of control and incentivized trait measures are not a function of the information available to subjects about the ideal job candidate.

²⁷ With respect to the correlation between IQ and openness, IQ is positively correlated with openness in the *Extroversion* treatment, negatively correlated with openness in the *Introversion* treatment, and essentially uncorrelated with openness in the *No Priming* treatment. One possibility is that the treatments inadvertently provide cues that openness is desired in the *Extroversion* treatment but not desired in the *Introversion* treatment; this does not appear to have been the case as we discuss in Section 7. Alternatively, subjects may arrive at the lab with beliefs about the correlations between desired traits. Specifically, subjects may believe that firms looking for extroversion (introversion) are also looking for openness (non-openness). In this case, more intelligent subjects who identify the extroversion cues would also manipulate their openness responses.

Table 3 are equal in the *Extroversion* and *Audit* treatments even after adjusting the p-values in Table 3 to account for multiple hypotheses.

For perspective, trimming the bottom half of subjects based on incentivized extroversion scores would eliminate 63 of the 84 subjects in the bottom half of the distribution of non-incentivized extroversions scores in the *Audit* treatment compared to 51 of 84 subjects in the *Extroversion* treatment. As an ordinal measure of personality traits (treating the non-incentivized measures as if they were the “true” personality traits), incentivized measures appear to work much better when subjects have reason to be apprehensive about misrepresentations.

In terms of correlations with other measures, the incentivized personality measures in the *Audit* treatment are essentially only correlated with the non-incentivized measure of the same trait. The lone exception is the negative correlation between IQ and incentivized conscientiousness in Column (4) of Table 5. We speculate that more intelligent individuals may appreciate that they tend to inflate their conscientiousness responses as prior studies have suggested that most “fakers” do (Viswesvaran and Ones [1999]) and worked to reign in their faking in the Evaluation session so as not to fail the audit—although it is unclear why this same logic would not apply to the correlation between IQ and incentivized extroversion in the *Audit* treatment. Alternatively, more intelligent subjects may have inferred (incorrectly) that conscientiousness items were most likely to be audited and reigned in their faking accordingly.

The negative correlation between IQ and incentivized conscientiousness contrasts with the positive correlation found in Vasilopoulos et al.’s [2005] studies of the effect of verification warnings. As discussed in Section 2, we suspect the difference in how IQ influences incentivized conscientiousness measures stems from the nature of the verification threat and how more intelligent individuals adapt their responses in the presence of different threats. Collectively,

however, our studies point to another role for intelligence in the incentivized measurement of personality when test-takers believe their responses are subject to verification.

Finally, we measured risk aversion because we hypothesized that risk averse individuals might reduce their misrepresentations in the *Audit* treatment when faced with the risk of an audit. In fact, we observe no significant relationships between risk aversion and incentivized personality scores in the *Audit* treatment.

6. Robustness of the experimental findings

6.1 Incentives and effort in the Baseline session

We assumed that all subjects supply the same effort in the Baseline session such that differences across the treatments are entirely attributable to differences in behavior in the Evaluation sessions. In reality, differences in intrinsic motivations in the absence of monetary incentives may lead to differences in the effort supplied in the Baseline sessions. If the sample is not balanced across treatments in terms of the “seriousness” with which subjects approached the Baseline session, then some effects attributed to the treatments may stem from differences in intrinsic motivation.

We construct two measures of “seriousness” to identify subjects who supplied low effort in the Baseline sessions to assess whether our findings are influenced by these subjects. First, we identify 42 subjects out of 642 (6.5%) who give the same response to every item on a screen more than once. Subjects were shown five Big 5 items on the screen at any time and had to respond to each item before moving on to a new screen with five more items. Though it is entirely possible that subjects could have the same response for five consecutive items, doing so

for several sets of items is more likely to reflect a lack of seriousness.²⁸ Second, we identify 42 subjects who provide identical responses for at least one of four pairs of items in which the statements are nearly complete opposites.²⁹

In short, excluding subjects using either measure of seriousness has no appreciable effects on either the changes in measured personality traits in Table 2 or the correlations in Table 3. Appendix B Table 3 reports estimates analogous to those in Table 4 excluding “unserious” subjects. “Unserious” subjects were relatively evenly distributed across treatments as evident from the sample sizes, and consequently the estimates are similar to those in Table 4. Likewise, the estimates using the other incentivized Big 5 measures as the dependent variables excluding “unserious” subjects are also similar to those in Table 5; these estimates as well as versions of Tables 2 and 3 excluding “unserious” subjects are available from the authors. This gives us confidence that the findings in Section 5 result from treatment effects in the Evaluation session rather than a lack of seriousness in the Baseline sessions.

6.2 Censored incentivized personality trait scores

The highest possible raw trait score on our personality test is 100, and the scatterplots in Figure 2 make it appear as though many subjects may have right-censored incentivized extroversion scores. In principle, this right-censoring could lead to inconsistent estimates of the slope coefficients in Table 4. In fact, right-censoring of the incentivized extroversion scores is fairly uncommon. The incentivized extroversion scores are right censored for only 6 observations in the *Extroversion* treatment, none in the *Introversion* treatment, and 1 observation

²⁸ One subject gave the same response to all 100 items, while 162 subjects gave the same answer for all items on a screen at least once, 19 did so at least three times, and 10 did so at least four times.

²⁹ The pairs are: “I wait for others to lead the way” and “I take charge,” “I leave my belongings around” and “I like order,” “I am not easily annoyed” and “I can be stirred up easily”, and “I inquire about others' feelings” and “I am indifferent to the feelings of others.”

each in the *No Priming* and *Audit* treatments. Nevertheless, we re-estimate the coefficients in Column (1) of Table 4 for the *Extroversion* treatment using the semi-parametric censored least absolute deviations (CLAD) estimator (Powell [1984]) given the heteroskedasticity evident in our data.³⁰ The estimates—reported in Appendix B Table 3—are similar to those in Table 4.

6.3 External validity with respect to the personality tests used by employers

The tests employers actually use to screen applicants are proprietary. As such, the DeYoung [2007] personality test might differ significantly from those used by employers. With this in mind, we compared our personality test to those of two major US retailers that require personality tests when individuals apply online. Of the 30 items on the first retailer’s test, 23% (7/30) of the items were identical to items on our personality test while 90% (27/30) were identical or nearly identical to items on our test.³¹ The second retailer’s test had 50 items, but many (27/50) of these were either not personality questions per se or were measuring personality traits other than the Big 5.³² On this test, only 2% (1/50) of the items were identical to items on our test, but 46% (23/50) of the items were nearly identical. The lower correspondence between items in this second test and ours reflects, to some extent, the intermingling of the personality items with other items. Overall, the similarities between the DeYoung test and those used by these retailers suggest that the DeYoung test is a reasonable approximation for the tests used by actual employers (at least among retailers).

³⁰ Likewise, right-censoring was uncommon for the other incentivized Big 5 measures with only 7 instances for conscientiousness and one instance each for agreeableness and openness. Because the 7 censored observations for conscientiousness are spread across treatments, this right-censoring has little effect on the estimates in Table 5.

³¹ We deem test items to be “nearly identical” to items on our test if they are simple negations of one of our items (e.g., “I am easily distracted” versus “I am not easily distracted”) or if the retailer’s wording was a straightforward rewording of one of our items (e.g., “I hate being the center of attention” versus “I am uncomfortable when I am the center of attention” or “I am the first to act” versus “I am the first to act at work”).

³² An example of the former is, “Based on your knowledge and experience, which of the following is the MOST important for providing quality customer service?” As examples of the latter, the retailer included three questions similar to items from our locus of control scale as well as the same optimism-pessimism scale that we use.

7. Do job ads indicate desired personality traits?

In the experiment, information about the ideal candidate's personality influences the correlation between incentivized personality measures and intelligence. The experimental job ads, however, may be extreme in the extent to which they prime extroversion and introversion relative to actual job ads. To our knowledge, no studies document the distribution of desired personality traits in a broad sample of job ads.³³

To fill this gap, we generate a list of adjectives associated with both extremes of each Big 5 personality trait by combining the adjective lists in Goldberg [1992] and Perugini and Leone [1996].³⁴ We then use the online *Roget's 21st Century Thesaurus, Third Edition*, to match each adjective in this seed list to synonyms. *Roget's* groups synonyms into four categories in terms of how close words are as synonyms; we use the first two “nearest” groups of synonyms to supplement our seed lists. The resulting adjective lists associated with each trait extreme are given in Appendix B Table 4.

We use Porter's [1980] stemming algorithm to reduce our adjective lists to stems or roots in order to identify words in ads regardless of whether they appear as adjectives, adverbs, or nouns. For instance, the adjective “audacious” (from the non-conscientiousness list) has the same stem (“audaci”) as the noun “audacity.” We then sum the number of distinct stems associated with each personality trait extreme appearing in a job ad.³⁵ This results in ten counts for each ad—one for each personality trait extreme.

³³ A number of studies document the personality demands in job ads for particular occupations (e.g., Headrick [2001] for chemists and Adkins [2004] for librarians).

³⁴ The Big 5 traits are a refinement of the “lexical hypothesis.” Under the lexical hypothesis, a language will develop adjectives that describe common, stable personality traits in order to efficiently convey specific traits. Closely related adjectives are likely to reflect the same personality trait, and the Big 5 emerged as five overarching, basic traits to which groups of adjectives belonged. Each Big 5 trait has facets that are associated with groups of common adjectives, and Goldberg [1992] and Perugini and Leone [1996] provide lists of these adjectives.

³⁵ Repeated use of a word or stem in an ad does not increase the count.

We apply this method to a sample of 142,618 job ads from Monster.com to assess how our experimental job ads compare to actual ads in the extent to which words associated with personality traits are represented. The sample consists of all job ads in the US posted to Monster.com over a two-week period in 2006.³⁶ No restrictions are placed on the sample; the ads include postings for jobs in a wide range of occupations and locations across the United States.³⁷

We note a few limitations of our approach. First, the adjective lists for each trait extreme contain different numbers of words depending on the number of words matched to a trait extreme.³⁸ Second, some words appear in ads for reasons that have nothing to do with personality traits. For instance, the stem “power” for the adjective “powerful” in the extroversion list appears in any job ad referencing “PowerPoint” as in our experimental job ads. In view of these shortcomings, we avoid comparisons across traits in terms of the counts within job ads.

Table 6 reports the mean number of words associated with each trait extreme as well as the 10th, 25th, 50th, 75th, and 90th percentiles of the distributions of these counts in the Monster.com sample. The table also reports the number of such words appearing the job ads used in the *Extroversion*, *Introversion*, and *No Priming* treatments. The means and distributions from the Monster.com job ad sample make it evident that many employers convey significant information about desired personalities in job ads. The job ads contain on average between 4 and 6 words associated extroversion, conscientiousness, agreeableness, emotional stability, and openness—

³⁶ Brencic and Norris [2012] use the same data to study how job ad characteristics such as education and work experience requirements—but not desired personality traits—are related to the duration of a vacancy’s posting on the job board.

³⁷ To facilitate comparisons with the experimental job ads, we analyze the text of the job ads in the “Job Description” field. Where employers did not adhere to standard ad formatting, we analyze the full text of the ad.

³⁸ The longest words lists are for extroversion (135), disagreeableness (123), and neuroticism (121), while the shortest are for conscientiousness (95) and openness (98). We exercised some discretion and removed words like “firm” and “experience” that appear in almost all job ads for reasons unrelated to personality, and likewise we replaced stems with full words in cases when the stem would likely appear in ads for reasons unrelated to personality. For instance, “accountable” in the conscientiousness list reduces to the stem “account.”

traits that are presumably desirable—and far fewer words associated with non-openness (0.21), non-conscientiousness (0.26), neuroticism (0.65), and disagreeableness (0.60).

Within the sample, job ads exhibit tremendous heterogeneity in the extent to which words associated with traits are represented—which is consistent with jobs and employers requiring different mixes of traits (just as with skills) in the production process. The number of words associated with a trait in jobs ads at the 90th (25th) percentile of the distribution is roughly double (half) the number of words associated with the trait in the median job ad for extroversion, conscientiousness, emotional stability, and openness. In terms of our experimental findings, this heterogeneity suggests that many job applicants take personality tests without a clear indication from the job descriptions of the ideal applicant’s personality, while many others take personality tests with an indication of what the employer is looking for in terms of personality.³⁹

Relative to the Monster.com job ads, our experimental job ads do not appear outlandish in the extent to which they primed particular traits. The *Extroversion* job ad contains 10 words associated with extroversion, which puts it at the 94th percentile of the Monster.com distribution. At the high end of the Monster.com distribution, the purpose of this job ad was to make it clear that the employer desired an extrovert, which the job ad clearly did without being completely over the top: the maximum number of words associated with extroversion in an ad in the Monster.com sample was 26. By contrast, the *Introversion* job ad contained only one word associated with introversion, which would put it between the 37th and 70th percentiles of the Monster.com distribution. Thus subjects in the experiment appear to have been able to detect the priming even with considerably more subtle hints than in the *Extroversion* treatment. Job ads that

³⁹ How much personality information is contained in job ads for positions requiring personality tests versus the information contained in job ads not requiring such tests remains an open question. Unfortunately, our data do not allow us to identify whether the vacancy required applicants to complete a personality test.

convey information about the desired personalities of applicants appear to be reasonably common in the labor market rather than an idiosyncratic feature of our experimental design.

Finally, we noted in Section 5 that the correlations between IQ and openness may result from inadvertent priming of openness in the *Extroversion* treatment and non-openness in the *Introversion* treatment. The evidence in Table 6, however, suggests that this was unlikely to have occurred as there are more words associated with openness in the *Introversion* treatment than in the *Extroversion* treatment, and neither experimental job ad contains any words associated with non-openness. If anything, the evidence in Table 6 suggests that the *Extroversion* treatment may have inadvertently primed agreeableness—largely because three words in the *Extroversion* job ad appear in both the extroversion and agreeableness adjective lists.

8. How Do Employers Use Personality Tests in Hiring?

What do employers aim to learn from personality tests? Are employers genuinely attempting to find workers with particular personality traits, or are they simply using such tests to aggregate useful information (i.e., personality and intelligence) to winnow the applicant pool? To study how employers use personality tests, we analyze the hiring practices of representative samples of British firms in the 2004 and 2011 waves of the Workplace Employment Relations Study (WERS). The WERS surveyed managers about whether the firm used personality or attitude tests and/or performance or competency tests to fill vacancies. Managers were also asked the channels used to recruit workers and the factors important when recruiting.

Table 7 reports the summary statistics concerning firms' hiring practices. Testing grew more common between 2004 and 2011. By 2011, 21 (47) percent of establishments used personality or attitude (performance or competency) tests when filling vacancies, while nearly 41 (68) percent of employees worked at establishments using personality or attitude (performance or

competency) tests. In 2004, firms were most likely to use personality or attitude tests when filling vacancies for sales and managerial positions and least likely to use these tests to fill vacancies for operators and routine unskilled laborers.⁴⁰ By contrast, firms were most likely to use performance or competency test to fill vacancies for professionals and technical occupations.

Among the factors that employers reported important in filling vacancies, most employers regarded qualifications or experience (90% in 2011) and skills (87%) to be important, but a similar fraction (81%) reported that applicants' motivation was an important factor. WERS did not explicitly ask about the importance of personality, but 2% of employers listed personality among the "other factors" they considered important.

Conditional on using a personality test in hiring, 79% of employers in 2011 also used a performance or competency test. That is, employers do not appear to substitute personality tests for performance tests despite the potential for such substitution evident from our experiment. Conditional on giving a performance or competency test, employers in 2004 were most likely to also use a personality test to fill vacancies for managers (50%) and sales workers (40%)—positions that require significant interaction with others necessitating a degree of extroversion.⁴¹

Finally, Table 8 reports the estimated marginal effects on the probability of using a personality or attitude test when filling vacancies of firm characteristics, hiring factors, and recruitment practices. Using the 2011 WERS employer sample, Column (1) reports estimates from a baseline specification with firm characteristics. Firms with more employees and a larger fraction of full-time employees were more likely to use personality tests—perhaps because the average cost of testing is lower for large firms and the stakes in hiring higher when workers are more likely to be filling full-time positions. The marginal effects of the factors important when

⁴⁰ Managers were asked whether these tests were used to fill vacancies in detailed occupations only in 2004.

⁴¹ These conditional proportions are based on the authors' calculations.

hiring added as controls in Column (2) are not statistically significant. Using the 2004 WERS employer sample in Column (3), however, firms indicating that an applicant's motivation was an important factor when hiring were an estimated 9.6 percentage points more likely to use a personality test when hiring, and firms indicating that the applicant's personality was important were 6.9 percentage points more likely to administer a personality test—though the latter estimate is not statistically significant at conventional levels.

Returning to the 2011 sample, Column (4) adds an indicator for whether the firm used performance or competency tests. Avoiding a causal interpretation given that this is surely endogenous, we simply note that firms using performance or competency tests were also more likely to use personality tests conditional on firm characteristics and the factors important when recruiting. Using a performance or competency test is associated with a 23.1 percentage point increase in the probability of using a personality or attitude test in hiring.

Columns (5) and (6) restrict the sample to firms that had hired in their largest occupational group in the last 12 months as only these firms were asked about the channels used for recruiting. Column (5) adds to the specification in Column (2) indicators for whether the firm recruited using the internet or print ads. Firms using the internet to recruit were 8.8 percentage points more likely to use a personality test in hiring, while firms using print ads were 6.8 percentage points less likely to use a personality test. Controlling for whether firms recruited via the internet in Column (6), using a performance or competency test in hiring was still associated with a 22.9 percentage point increase in the probability of using a personality test in hiring.

To summarize, we find that a large and increasing fraction of employers in Great Britain used personality tests to screen applicants when filling vacancies between 2004 and 2011—particularly for managers and salespersons. The use of personality testing for these occupations

undoubtedly facilitates the sorting of gregarious workers into occupations requiring social interactions documented in Krueger and Schkade [2008]. While most employers indicate that they prioritize skills, qualifications and experience, nearly 4 out of 5 employers indicate that they look for motivated workers, while still others indicate that a worker's personality is important. Consistent with the expressed demand for personality traits in job ads, we find that employers who give personality tests are also likely to give performance or competency tests—suggesting that employers do not use personality tests as a substitute for performance or competency tests despite the potential for such substitution evident in our experiment.

9. Conclusion

We examine the effects on the measurement of personality of incentives, information, and situations (here whether misrepresentations may be detected and punished) in a within-subject laboratory experiment with incentives (but not directives) for misrepresentation. Incentives for misrepresentation have dramatic effects with incentivized and non-incentivized personality measures being only weakly to moderately correlated and the incentivized means being a full standard deviation higher than non-incentivized means for most traits. The distortionary effects of incentives, however, are largely mitigated in our experiment when subjects know that their responses may be verified with misrepresentations rendering them ineligible for the bonus. The correlations between incentivized and non-incentivized personality traits in our *Audit* treatment range between 0.6 and 0.8—not far from the *Test-Retest* correlations of around 0.9.

By contrast, we find that the availability of information about desired personality traits—which we show is a fairly common feature of actual job ads—has little or no effect on the correlations between incentivized and non-incentivized personality measures. Information about desired personality traits, however, does appear to generate correlations between IQ and the

incentivized measure of the personality trait being sought in the job ad. Consistent with existing evidence in the psychology literature, we also find some evidence that IQ may influence incentivized conscientiousness measures when subjects know that misrepresentations may be detected and punished. Taken together, these findings suggest multiple channels through which cognition and intelligence may influence incentivized personality measures.

Among other traits, we find evidence that subjects with an internal locus of control misrepresent themselves more for several traits across treatments. We speculate that internal individuals may believe the returns to the effort they put into faking to be higher than less internal subjects given that the returns to faking on any given item are unclear in our experiment.

A number of important unresolved questions follow from our study with practical implications for hiring in the labor market. First, are incentivized personality measures similar to non-incentivized measures—or at least less influenced by incentives—using alternative verification warnings? Our within-subject laboratory experiment is particularly useful in establishing the relationship between incentivized and non-incentivized measures in the presence of the audit, but the audit procedure itself is unrealistic for actual firms. The descriptive evidence from British firms that we present suggests that firms are genuinely interested in using personality tests not to substitute for other assessments but to learn about personality. Assuming that firms desire personality tests to have more discriminatory power, identifying optimal and implementable verification warnings remains an important area for research. Unfortunately, the verification methods available to employers (e.g., contacting references) are logistically challenging in laboratory experiments and thus not likely to be credible and possibly even deceptive. As such, this issue likely will have to be investigated with field experiments.

Second, do the links between IQ and incentivized personality measures affect estimates of the criterion validity of incentivized personality measures relative to job performance? Our study establishes conditions—namely the availability of information about desired traits and the presence of verification warnings—under which incentivized personality measures may be correlated with IQ. If IQ is a significant predictor of job performance, then estimates of the criterion validity of incentivized personality measures may vary according to these conditions.

Finally, to what extent do our findings extend to other environments in which personalities are assessed in the presence of incentives for misrepresentation such as cover letters, job interviews or LinkedIn profiles? Can individuals misrepresent their personalities just as easily in these other venues as in personality tests, or are the signals regarding personality in these settings more indicative of actual personality? Likewise, do more intelligent individuals perform better in interviews because they are better equipped to infer what interviewers want to hear, or does an in-person screening tool provide a useful aggregation of personality and other traits such as emotional intelligence?

References

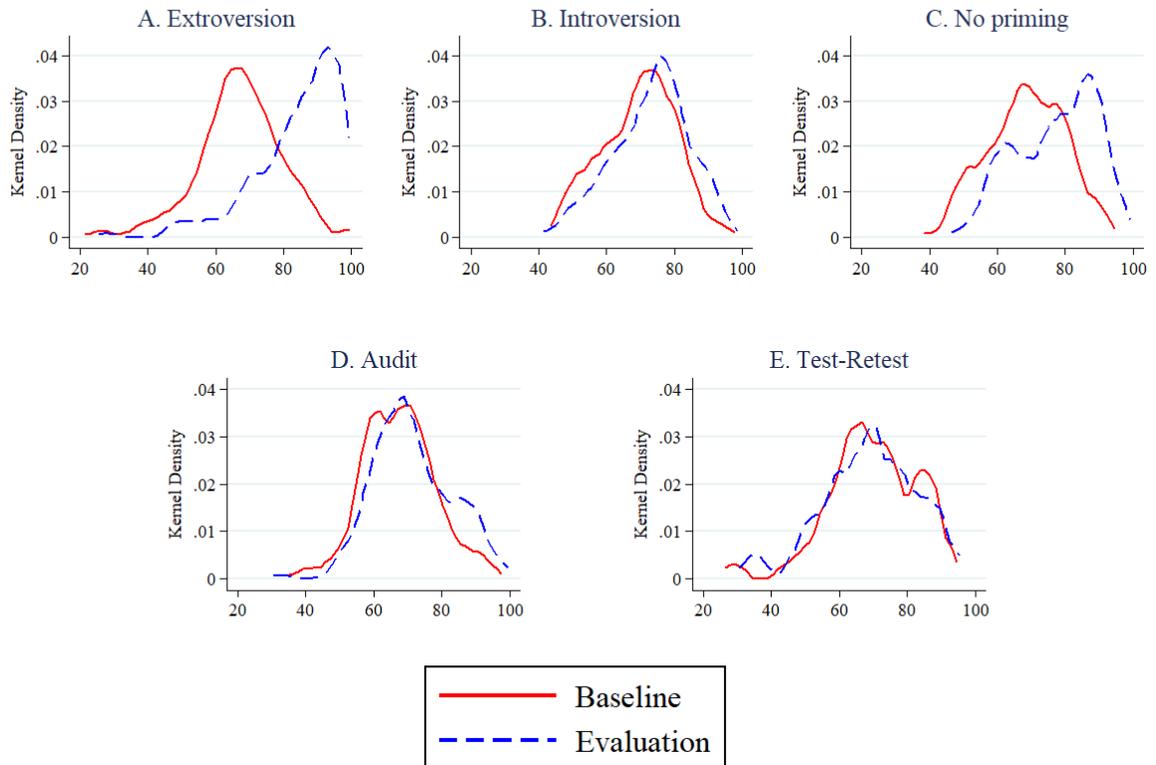
- Adkins, Denice.** 2004. "Changes in public library youth services: A content analysis of youth services job advertisements." *Public Library Quarterly*, 23(3-4), 59-73.
- Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz.** 2011. "Personality Psychology and Economics." *Handbook of the Economics of Education*, 4, 1-181.
- Anderson, Michael.** 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, 103(484), 1481-1495.
- Autor, David H., and David Scarborough.** 2008. "Does job testing harm minority workers? Evidence from retail establishments." *The Quarterly Journal of Economics*, 123 (1), 219-277.
- Barrick, Murray R., and Michael K. Mount.** 1991. "The big five personality dimensions and job performance: a meta-analysis." *Personnel Psychology*, 44 (1), 1-26.
- Becker, Thomas and Alan Colquitt.** 1992. "Potential versus actual faking of a biodata form: analysis along several dimensions of item type." *Personnel Psychology*, 45, 389-406.
- Benjamini, Yoav, and Yosef Hochberg.** 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289-300.
- Birkeland, Scott A., Todd M. Manson, Jennifer L. Kisamore, Michael T. Brannick, and Mark A. Smith.** 2006. "A meta-analytic investigation of job applicant faking on personality measures." *International Journal of Selection and Assessment*, 14 (4), 317-335.
- Borghans, Lex, Golsteyn, Bart, Heckman, James, and Humphries, John Eric.** 2011. "Identification problems in personality psychology." *Personality and Individual Differences*, 51(3), 315-320.
- Brencic, Vera and John B. Norris.** 2012. "Employers' Online Recruitment and Screening Practices." *Economic Inquiry*, 50(1), 94-111.
- Cattell, Raymond B.** 1963 "Theory of fluid and crystallized intelligence: A critical experiment." *Journal of Educational Psychology*, 54(1), 1.
- Christiansen, Neil, Richard Goffin, Norman Johnston, and Mitchell Rothstein.** 1994. "Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions." *Personnel Psychology*, 47, 847-860.
- Dal Bo, Pedro, Frederico Finan, and Martin Rossi.** 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." *Quarterly Journal of Economics*, 128(3), 1169-1218.
- Davidson, Russell and James MacKinnon.** 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- DeYoung, Colin G., Lena C. Quilty, and Jordan B. Peterson.** 2007. "Between facets and domains: 10 aspects of the Big Five." *Journal of Personality and Social Psychology*, 93(5), 880.
- Donovan, John, Stephen Dwight, and Gregory Hurtz.** 2003. "An assessment of the prevalence, severity and verifiability of entry-level applicant faking using the randomized response technique." *Human Performance*, 16, 81-106

- Duckworth, A.L., Quinn, P.D., Lynam, D.R., Loeber, R. and Stouthamer-Loeber, M.** 2011. "Role of test motivation in intelligence testing." *Proceedings of the National Academy of Sciences*, 108(19), 7716-7720.
- Dwight, Stephen A. and John J. Donovan.** 2003. "Do Warnings Not to Fake Reduce Faking?" *Human Performance*, 16(1), 1-23.
- Fischbacher, Urs.** 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental economics*, 10(2), 171-178.
- Freyd, Max.** 1926. "Selection of promotion salesmen." *Journal of Personality. Research*, 5, 142-156.
- Furnham, Adrian.** 1990 "Faking personality questionnaires: Fabricating different profiles for different purposes." *Current Psychology*, 9(1), 46-55.
- Furnham, Adrian.** 1997. "Knowing and Faking One's Five-Factor Personality Score." *Journal of Personality Assessment*, 69(1), 229-243.
- Ghiselli, Edwin E., and Richard P. Barthol.** 1953. "The validity of personality inventories in the selection of employees." *Journal of Applied Psychology*, 37 (1), 18.
- Goffin, Richard D., and Neil D. Christiansen.** 2003 "Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers." *International Journal of Selection and Assessment*, 11(4), 340-344.
- Goldberg, Lewis R.** 1981. "Language and individual differences: The search for universals in personality lexicons." *Review of Personality and Social Psychology*, 2 (1), 141-165.
- Goldberg, Lewis R.** 1992. "The development of markers for the Big-Five factor structure." *Psychological Assessment*, 4, 26-42.
- Headrick, Kurt L.** 2001. "Want ads, job skills, and curriculum: A survey of 1998 chemistry help-wanted ads." *Journal of Chemical Education*, 78(9), 1281.
- Heckman, James J., and Tim Kautz.** 2012. "Hard evidence on soft skills." *Labour Economics*, 19 (4), 451-464.
- Hoffman, Mitchell, Lisa B. Kahn, and Danielle Li.** 2018. "Discretion in hiring." *Quarterly Journal of Economics*, 133(2), 765-800.
- Holden, Ronald and Nicole Hibbs.** 1995. "Incremental validity of response latencies for detecting fakers on a personality test." *Journal of Research in Personality*, 29, 362-372.
- Holt, Charles A. and Susan K. Laury.** 2002. "Risk Aversion and Incentive Effects." *American Economic Review*, 92(5), 1644-1655.
- Hough, Leaetta, Newell Eaton, Marvin Dunnette, John Kamp, and Rodney McCloy.** 1990. "Criterion-related validities of personality constructs and the effect of response distortion on those validities." *Journal of Applied Psychology*, 75(5), 581-595.
- Hough, Leaetta M., and Frederick L. Oswald.** 2000. "Personnel selection: Looking toward the future--Remembering the past." *Annual Review of Psychology*, 51(1), 631-664.
- Hurtz, Gregory M., and John J. Donovan.** 2000. "Personality and job performance: The Big Five revisited." *Journal of Applied Psychology*, 85 (6), 869-879.
- Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi.** 1997. "The effects of human resource management practices on productivity: A study of steel finishing lines." *The American Economic Review*, 87 (3) , 291-313.
- Jackson, Douglas N.** 1994. *Jackson Personality Inventory-Revised manual*. Port Huron, MI: Sigma Assessment Systems.

- Kleinmann, Martin, Pia V. Ingold, Filip Lievens, Anne Jansen, Klaus G. Melchers and Cornelius J. König.** 2011. "A different look at why selection procedures work : The role of candidates' ability to identify criteria." *Organizational Psychology Review*, 1(2), 128–146.
- Kroger, Rolf and Linda Wood.** 1993. "Reification, "faking," and the Big Five." *American Psychologist*, 48, 1297-1298.
- Krueger, Alan B. and David Schkade.** 2008. "Sorting in the Labor Market: Do Gregarious Workers Flock to Interactive Jobs?" *Journal of Human Resources*, 43(4), 859-883.
- McGee, Andrew and Peter McGee.** 2016. "Search, effort, and locus of control." *Journal of Economic Behavior & Organization*, 126, 89-101.
- McLeod, Beverly A., and Randy L. Genereux.** 2008. "Predicting the acceptability and likelihood of lying: The interaction of personality with type of lie." *Personality and Individual Differences*, 45 (7), 591-596.
- McCrae, Robert R., and Oliver P. John.** 1992. "An introduction to the five-factor model and its applications." *Journal of Personality*, 60 (2), 175-215.
- McFarland, Lynn A., and Ann Marie Ryan.** 2000. "Variance in faking across noncognitive measures." *Journal of Applied Psychology*, 85 (5), 812.
- Morgeson, Frederick, Michael Campion, Robert Dipboye, John Hollenbeck, Kevin Murphy, and Neal Schmitt.** 2007a. "Reconsidering the use of personality tests in personnel selection contexts." *Personnel Psychology*, 60 (3), 683-729
- Morgeson, Frederick P., Michael A. Campion, Robert L. Dipboye, John R. Hollenbeck, Kevin Murphy, and Neal Schmitt.** 2007b. "Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection." *Personnel Psychology*, 60 (4), 1029-1049.
- Ones, Deniz and Chockalingam Viswesvaran.** 1998. "The effects of social desirability and faking on personality and integrity assessment for personnel selection." *Human Performance*, 11, 245–269.
- Ones, Deniz S., Stephan Dilchert, Chockalingam Viswesvaran, and Timothy A. Judge.** 2007. "In support of personality assessment in organizational settings." *Personnel Psychology*, 60 (4), 995-1027.
- Paulhus, Delroy L.** 1984. "Two-component models of socially desirable responding." *Journal of Personality and Social Psychology*, 46 (3), 598-609.
- Paulhus, Delroy L.** 2002. "Socially desirable responding: The evolution of a construct." *The role of constructs in psychological and educational measurement* 4969.
- Perugini, Marco, and Luigi Leone.** 1996. "Construction and validation of a Short Adjectives Checklist to measure Big Five (SACBIF)." *European Journal of Psychological Assessment*, 12(1), 33-42.
- Porter, Martin.** 1980. "An algorithm for suffix stripping." *Program*, 14(3), 130–137.
- Powell, James L.** 1984. "Least Squares Absolute Deviations Estimation for the Censored Regression Model." *Journal of Econometrics*, 25, 303-325.
- Raven, John C.** 1998. *Raven's progressive matrices*. Oxford: Oxford Psychologists Press.
- Roberts, Brent W.** 2009. "Back to the future: Personality and assessment and personality development." *Journal of Research in Personality*, 43(2), 137-145.

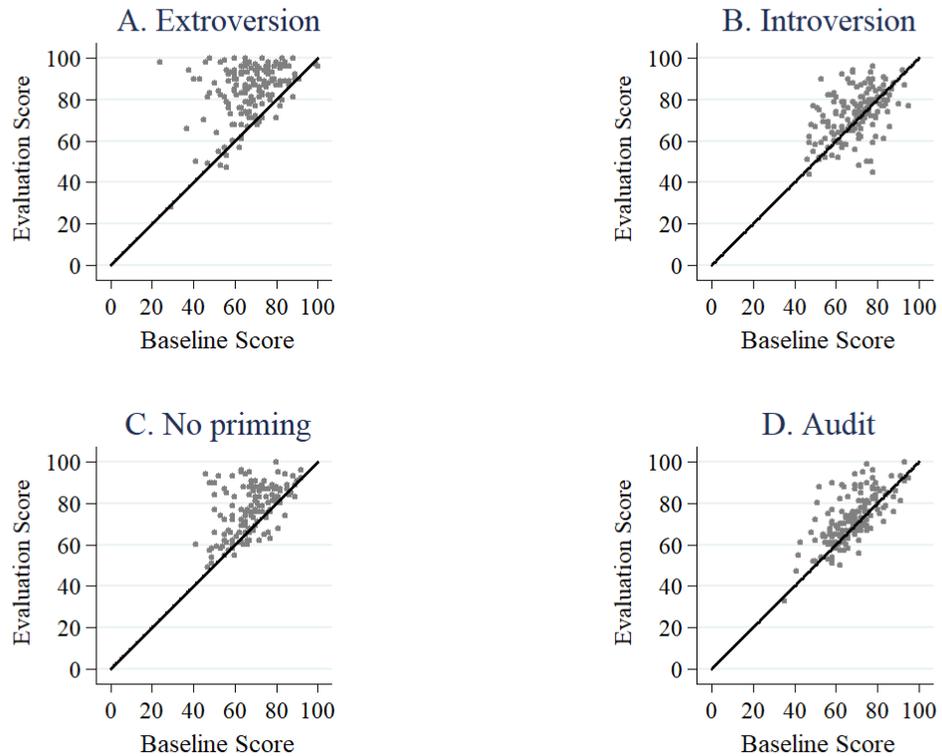
- Rosse, Joseph G., Mary D. Stecher, Janice L. Miller, and Robert A. Levin.** 1998. "The impact of response distortion on preemployment personality testing and hiring decisions." *Journal of Applied Psychology*, 83 (4), 634.
- Rothstein, Mitchell G. and Goffin, Richard D.** 2006. "The use of personality measures in personnel selection: What does current research support?" *Human Resource Management Review*, 16(2), 155-180.
- Rotter, Julian B.** 1966. "Generalized Expectancies for Internal Versus External Control of Reinforcement." *Psychological Monographs General and Applied*, 80 (1, Whole No. 609).
- Scheier, Michael F., Charles S. Carver, and Michael W. Bridges.** 1994. "Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test." *Journal of Personality and Social Psychology* 67 (6), 1063.
- Schmidt, Frank L., and John Hunter.** 1998. "The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings." *Psychological Bulletin*, 124, 262-274.
- Segal, Carmit.** 2012. "Working when no one is watching: Motivation, test scores, and economic success." *Management Science*, 58 (8), 1438-1457.
- Snyder, Mark.** 1974. "Self-monitoring of expressive behavior." *Journal of Personality and Social Psychology*, 30, 526-537.
- Tett, Robert P., Kurt A. Freund, Neil D. Christiansen, Kevin E. Fox, and John Coaster.** 2012. "Faking on self-report emotional intelligence and personality tests: Effects of faking opportunity, cognitive ability, and job type." *Personality and Individual Differences*, 52(2), 195-201.
- Vasilopoulos, Nicholas L., Jeffrey M. Cucina, and Julia M. McElreath.** 2005. "Do warnings of response verification moderate the relationship between personality and cognitive ability?" *Journal of Applied Psychology*, 90(2), 306-322.
- Velicer, Wayne and Barbara Weiner.** 1975. "Effects of sophistication and faking sets on the Eysenck Personality Inventory." *Psychological Reports*, 37, 71-73.
- Viswesvaran, Chockalingam and Deniz Ones.** 1999. "Meta-analyses of fakability estimates: implications for personality measurement." *Educational and Psychological Measurement*, 59 (2), 197-210.
- Weber, Lauren, and Elizabeth Dwoskin.** 2014. "Are workplace personality tests fair? Growing use of tests sparks scrutiny amid questions of effectiveness and workplace discrimination." *The Wall Street Journal*, September 29, 2014.

Figure 1: Extroversion kernel densities



Notes: Kernel density plots of raw extroversion scores from the Baseline (non-incentivized) and Evaluation (incentivized) sessions by treatment are displayed. Panel A depicts the densities for the *Extroversion* treatment, Panel B the *Introversion* treatment, Panel C the *No Priming* treatment, Panel D the *Audit* treatment and Panel E the *Test-Retest* treatment. The Evaluation session in the *Test-Retest* treatment was the same as the Baseline session and thus not incentivized. The p-values for the two-sample Kolmogorov-Smirnov tests for the equality of the distributions between the Baseline and Evaluation sessions are 0.000, 0.029, 0.000, and 0.065 for Panels A, B, C, and D, respectively. The estimated probability that a random draw from the non-incentivized score distribution is larger than a random draw from the incentivized score distribution is 0.154 in Panel A, 0.407 in Panel B, 0.300 in Panel C, and 0.406 in Panel D.

Figure 2: Extroversion scatter plots



Notes: The figure displays scatterplots of the raw, Baseline session (non-incentivized) extroversion score on the x-axis against the raw, Evaluation session (incentivized) extroversion score on the y-axis by treatment. The plots in each panel also depict a 45-degree line for use as a reference. Panel A depicts the scatterplot for the *Extroversion* treatment, Panel B the *Introversion* treatment, Panel C the *No Priming* treatment, and Panel D the *Audit* treatment. The Evaluation session in the *Test-Retest* treatment was the same as the Baseline session and thus not incentivized.

Table 1: Summary statistics for measures other than personality

	Treatment				KW p-value (5)
	Extroversion (1)	Introversion (2)	No Priming (3)	Audit (4)	
A. Measured in the Baseline Session					
Male	0.44	0.49	0.51	0.48	0.667
Age	21.91 (4.52)	20.82 (3.12)	20.46 (2.39)	21.11 (4.07)	0.007
White	0.19	0.29	0.24	0.21	0.157
Self-deception	5.73 (3.47)	6.23 (3.60)	5.37 (3.08)	5.90 (3.44)	0.252
Impression management	6.34 (3.83)	5.82 (3.46)	5.86 (3.38)	5.99 (3.82)	0.704
Self-monitoring	27.34 (8.33)	28.88 (8.67)	28.34 (7.96)	28.42 (8.38)	0.436
Machiavellianism	18.32 (4.65)	19.34 (4.64)	18.84 (4.61)	18.14 (4.88)	0.107
Acceptability of lies for personal gain	3.60 (1.61)	3.35 (1.44)	3.60 (1.35)	3.43 (1.46)	0.313
Acceptability of lies to be kind	5.01 (1.46)	4.97 (1.32)	5.17 (1.35)	4.92 (1.42)	0.334
Acceptability of lies to avoid conflict	4.78 (1.62)	4.59 (1.43)	4.79 (1.36)	4.72 (1.24)	0.382
Acceptability of lies to gain social acceptance	4.13 (1.69)	3.84 (1.52)	4.20 (1.53)	3.96 (1.49)	0.124
Optimism	20.93 (4.65)	21.24 (4.86)	20.62 (4.86)	20.26 (4.52)	0.139
B. Measured in the Evaluation Session					
IQ	50.12 (5.83)	50.27 (4.73)	50.75 (4.79)	49.20 (5.27)	0.057
Risk aversion	5.64 (1.98)	5.41 (2.02)	5.33 (2.12)	5.86 (2.03)	0.155
Self-deception	7.08 (4.72)	6.51 (4.19)	6.44 (3.83)	6.31 (4.12)	0.722
Impression management	6.51 (4.24)	5.90 (3.91)	6.13 (3.95)	5.28 (3.75)	0.054
Optimism	22.61 (5.00)	22.26 (4.81)	21.93 (4.61)	21.16 (4.50)	0.012
Locus of control	11.65 (2.15)	11.63 (1.94)	11.72 (2.01)	11.67 (1.95)	0.938
# of personality tests taken in job applications	0.94 (1.35)	0.89 (1.37)	0.76 (1.27)	0.96 (1.36)	0.321
# of subjects	167	160	147	168	

Notes: Standard deviations in parentheses. Column (5) reports the p-values for the rank-based nonparametric Kruskal-Wallis test for significant differences across treatment for each variable.

Table 2: Personality trait scores with and without incentives

	Extroversion	Introversion	Treatment No Priming	Audit	KW p-value
	(1)	(2)	(3)	(4)	(5)
			<u>Extroversion</u>		
Baseline	67.04 (12.33)	69.47 (10.89)	68.56 (11.31)	67.38 (10.68)	0.137
Evaluation	84.19 (12.76)	73.07 (11.24)	77.30 (11.88)	71.38 (11.71)	0.000
			<u>Conscientiousness</u>		
Baseline	67.76 (11.11)	66.42 (10.32)	65.03 (10.71)	66.45 (11.02)	0.226
Evaluation	82.05 (10.79)	81.96 (11.84)	78.06 (12.53)	70.89 (11.93)	0.000
			<u>Agreeableness</u>		
Baseline	74.12 (9.12)	74.86 (9.69)	73.80 (9.26)	74.32 (9.92)	0.505
Evaluation	80.06 (8.46)	81.36 (9.35)	78.27 (8.86)	75.64 (9.73)	0.000
			<u>Openness/Intellect</u>		
Baseline	70.95 (9.98)	73.34 (8.96)	72.03 (9.52)	71.98 (9.40)	0.111
Evaluation	82.98 (9.99)	78.51 (8.13)	79.21 (9.79)	73.02 (10.54)	0.000
			<u>Neuroticism</u>		
Baseline	58.35 (12.77)	57.26 (12.42)	59.36 (13.16)	59.09 (12.55)	0.373
Evaluation	37.27 (12.72)	41.73 (13.36)	43.71 (14.76)	52.62 (13.85)	0.000
# of subjects	167	160	147	168	

Notes: Each panel reports the mean non-incentivized (Baseline) and incentivized (Evaluation) scores for each personality trait by treatment. Columns (1) to (4) report the means for the *Extroversion*, *Introversion*, *No priming*, and *Audit* treatments, respectively. Column (5) reports the p-values for the rank-based nonparametric Kruskal-Wallis test for significant differences across treatments for the variable in the far-left column. The p-values for the nonparametric Wilcoxon signed-rank tests of equality of the means for the Baseline and Evaluation scores for each trait in every treatment are less than 0.001 with the exception of openness/intellect (0.337) and agreeableness (0.192) in the *Audit* treatment. Standard deviations are given in parentheses.

Table 3: Correlations between trait scores with & without incentives

	Treatment				
	Extroversion	Introversion	No Priming	Audit	Test-Retest
	(1)	(2)	(3)	(4)	(5)
<u>A. Pearson correlation coefficient</u>					
Extroversion	0.38 [0.203,0.796]	0.45 [0.557,0.973]	0.50	0.69 [0.001,0.014]	0.94
Conscientiousness	0.35 [0.454,0.973]	0.21 [0.036,0.330]	0.43	0.59 [0.019,0.209]	0.93
Agreeableness	0.55 [0.523,0.973]	0.53 [0.659,0.973]	0.49	0.71 [0.068,0.503]	0.92
Openness	0.42 [0.108,0.643]	0.63 [0.499,0.973]	0.57	0.78 [0.000,0.001]	0.83
Neuroticism	0.30 [0.164,0.760]	0.45 [0.927,0.973]	0.44	0.64 [0.000,0.006]	0.94
<u>B. Spearman's rank correlation coefficient</u>					
Extroversion	0.31	0.44	0.47	0.68	0.93
Conscientiousness	0.32	0.15	0.33	0.62	0.94
Agreeableness	0.54	0.44	0.55	0.73	0.90
Openness	0.39	0.64	0.48	0.76	0.77
Neuroticism	0.24	0.43	0.35	0.63	0.93
<u>C. Kendall's tau-b coefficient</u>					
Extroversion	0.23	0.32	0.37	0.51	0.80
Conscientiousness	0.24	0.10	0.24	0.48	0.79
Agreeableness	0.40	0.33	0.43	0.58	0.78
Openness	0.28	0.46	0.36	0.61	0.60
Neuroticism	0.17	0.32	0.27	0.49	0.78
# of subjects	167	160	147	168	45

Notes: Panel A reports pairwise Pearson correlation coefficients for the non-incentivized and incentivized trait scores measured in the Baseline and Evaluation sessions, respectively, by treatment, while Panels B and C report the analogous Spearman rank correlation coefficients and Kendall's tau-b coefficients, respectively. All of the correlations are statistically significant at the 1% level. Panel A also reports in brackets p-values for tests involving the Pearson correlation coefficients. Columns (1) and (2) report p-values for tests of the equality of the correlation in that column with the *No Priming* correlation in Column (3), while Column (4) reports p-values for a test of the equality of the *Audit* correlation with the *Extroversion* correlation in Column (1). The first p-value in brackets is an unadjusted p-value estimated by regressing standardized, incentivized measures on the standardized, non-incentivized measure along with treatment indicators and interactions between the treatment indicators and the non-incentivized measure, while the second p-value in brackets is the Holm-Sidak multiple hypothesis corrected p-value for the 15 hypotheses tested in the panel.

Table 4: Regressing incentivized extroversion on subject characteristics

	Treatment				p-value of (1)=...=(3)=0 (5)
	Extroversion (1)	Introversion (2)	No Priming (3)	Audit (4)	
Non-incentivized extroversion	0.413*** (0.135)	0.483*** (0.115)	0.426*** (0.114)	0.632*** (0.085)	0.000
IQ	0.196*** (0.065)	-0.178*** (0.068)	0.044 (0.081)	-0.004 (0.064)	0.001
Risk aversion	0.153** (0.068)	0.087 (0.090)	0.019 (0.072)	-0.028 (0.055)	0.110
Locus of control	-0.065 (0.078)	0.045 (0.075)	0.112 (0.080)	0.044 (0.062)	0.392
Acceptability of lies for gain	0.032 (0.103)	0.047 (0.086)	-0.012 (0.110)	0.057 (0.077)	0.938
Self-deception	0.076 (0.083)	0.036 (0.094)	-0.048 (0.100)	0.092 (0.071)	0.745
Impression management	-0.149 (0.106)	0.090 (0.102)	0.022 (0.110)	-0.069 (0.073)	0.424
Self-monitoring	-0.109 (0.089)	-0.070 (0.122)	0.161* (0.089)	0.062 (0.088)	0.167
Machiavellianism	0.027 (0.098)	-0.008 (0.124)	-0.028 (0.093)	0.011 (0.095)	0.982
Optimism	0.087 (0.089)	-0.033 (0.089)	0.000 (0.095)	-0.100 (0.081)	0.780
Constant	-0.000 (0.072)	0.000 (0.073)	0.000 (0.076)	0.000 (0.059)	
R ²	0.237	0.264	0.285	0.499	

Notes: Columns (1) to (4) report coefficient estimates from separate OLS regressions for each treatment. The dependent variable in each regression is the incentivized extroversion score from the Evaluation session standardized using the mean and standard deviation of the incentivized scores from subjects in the same treatment. The non-incentivized extroversion score used in the controls is standardized using the mean and standard deviation for the non-incentivized scores from subjects in the same treatment. The remaining controls—all of which were measured in the Baseline session except for locus of control—are standardized using the full sample of all subjects in all treatments. The standard errors reported in parentheses are scaled by the square of the observation's variance estimate (i.e., the diagonal element of the projection matrix) as suggested by Davidson and MacKinnon [1993] and implemented in Stata using the “vce(hc3)” option. Column (5) reports the p-value for a heteroscedasticity-robust Wald test for the hypotheses that the coefficients in Columns (1) to (3) are jointly significant when estimated in a fully-interacted model. Significance levels: *** p<0.01, ** p<0.05, * p<0.10

Table 5: Regressing other incentivized traits on subject characteristics

	Treatment				p-value of (1)=(2)=(3)=0 (5)
	Extroversion (1)	Introversion (2)	No Priming (3)	Audit (4)	
A. Conscientiousness					
Non-incentivized score	0.365*** (0.113)	0.183* (0.109)	0.396*** (0.104)	0.525*** (0.097)	0.000
IQ	0.074 (0.077)	-0.042 (0.091)	0.004 (0.077)	-0.121* (0.072)	0.765
Risk aversion	0.101 (0.089)	-0.025 (0.088)	0.026 (0.073)	-0.004 (0.066)	0.685
Locus of control	-0.001 (0.084)	0.254*** (0.074)	0.177** (0.080)	-0.022 (0.067)	0.001
B. Agreeableness					
Non-incentivized score	0.567*** (0.078)	0.714*** (0.112)	0.484*** (0.161)	0.762*** (0.081)	0.000
IQ	-0.036 (0.067)	0.069 (0.070)	0.127* (0.072)	0.034 (0.052)	0.232
Risk aversion	0.109 (0.077)	0.004 (0.063)	0.176** (0.070)	-0.053 (0.059)	0.040
Locus of control	0.033 (0.074)	0.110 (0.071)	0.101 (0.064)	0.079 (0.056)	0.168
C. Openness/Intellect					
Non-incentivized score	0.427*** (0.105)	0.602*** (0.080)	0.575*** (0.100)	0.768*** (0.061)	0.000
IQ	0.163** (0.071)	-0.090 (0.066)	0.041 (0.069)	0.011 (0.065)	0.060
Risk aversion	0.124 (0.077)	0.029 (0.079)	0.026 (0.062)	0.061 (0.051)	0.404
Locus of control	-0.063 (0.073)	0.163** (0.077)	0.044 (0.073)	-0.022 (0.049)	0.134
D. Neuroticism (emotional stability)					
Non-incentivized score	0.243** (0.107)	0.426*** (0.084)	0.432*** (0.111)	0.669*** (0.084)	0.000
IQ	-0.143** (0.072)	0.040 (0.070)	-0.047 (0.076)	-0.060 (0.062)	0.201
Risk aversion	-0.121 (0.086)	-0.070 (0.086)	-0.047 (0.073)	0.012 (0.058)	0.387
Locus of control	0.018 (0.081)	-0.231*** (0.072)	-0.131 (0.083)	0.015 (0.068)	0.006

Notes: Each column reports coefficient estimates from OLS regressions similar to those in Table 4 by treatment. The incentivized trait score for the trait specified in the panel is regressed on the non-incentivized trait score and the same controls used in Table 4. The complete estimates are reported in Appendix B Table 2. The standard errors are again those recommended in Davidson and MacKinnon [1993]. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 6: Frequency distributions of trait-related adjectives in job ads

Trait extreme	Monster.com Job Ad Distribution						Job descriptions in experiment		
	Mean (s.d.) (1)	10 th (2)	25 th (3)	50 th (4)	75 th (5)	90 th (6)	Extroversion (7)	Introversion (8)	No Priming (9)
Extroversion	4.50 (3.00)	1	2	4	6	9	10	4	2
Introversion	1.05 (1.05)	0	0	1	2	2	0	1	0
Conscientiousness	5.92 (3.43)	2	3	6	8	11	1	1	1
Non-conscientiousness	0.26 (0.52)	0	0	0	0	1	0	0	0
Emotional stability	4.39 (2.86)	1	2	4	6	8	4	3	1
Neuroticism	0.65 (0.88)	0	0	0	1	2	1	0	0
Agreeableness	3.46 (2.41)	1	2	3	5	7	6	3	2
Disagreeableness	0.60 (0.76)	0	0	0	1	2	0	0	0
Openness	4.73 (3.15)	1	2	4	7	9	2	3	1
Non-openness	0.21 (0.49)	0	0	0	0	1	0	0	0

Notes: The first six columns report summary statistics for the distributions of the number of words associated with trait extremes in the sample of 142,618 job ads from Monster.com. The last three columns report the number of words associated with trait extremes appearing in the job ads used in the experiment.

Table 7: Summary statistics for WERS 2004 and 2011 employer data

Proportion	Establishment weights		Employee weights	
	2004 (1)	2011 (2)	2004 (3)	2011 (4)
Using personality or attitude test	0.17	0.21	0.36	0.41
Using personality or attitude test to fill vacancies in specified occupation:				
Managerial	0.11	0.17	0.29	0.38
Non-managerial	0.13	0.13	0.25	0.20
Professionals	0.09	.	0.24	.
Technical	0.10	.	0.18	.
Administrative	0.05	.	0.11	.
Skilled trades	0.08	.	0.10	.
Personal service	0.07	.	0.04	.
Sales	0.13	.	0.20	.
Operators	0.04	.	0.08	.
Routine unskilled	0.05	.	0.07	.
Using performance or competency test	0.43	0.47	0.60	0.68
Using performance or competency test to fill vacancies in specified occupation:				
Managerial	0.18	0.31	0.31	0.54
Non-managerial	0.40	0.40	0.55	0.59
Professionals	0.31	.	0.39	.
Technical	0.29	.	0.33	.
Administrative	0.27	.	0.39	.
Skilled trades	0.24	.	0.26	.
Personal service	0.26	.	0.22	.
Sales	0.22	.	0.27	.
Operators	0.21	.	0.29	.
Routine unskilled	0.12	.	0.17	.
Recruiting channels used when filling vacancies:				
Internet (any)	0.05	0.44	0.08	0.63
Print ad	0.61	0.32	0.75	0.37
Factors important when recruiting new employees:				
References	0.70	0.66	0.72	0.70
Skills	0.82	0.87	0.88	0.92
Referrals	0.41	0.41	0.31	0.33
Qualifications or experience	0.88	0.90	0.89	0.92
Motivation	0.81	0.81	0.82	0.83
Personality	0.04	0.02	0.03	0.02
Percentage of employees full-time	0.67	0.68	0.72	0.73
Number of employees	31.94	31.84	516.47	709.49

Table 8: Probit models of the probability of using personality tests in hiring

Controls	(1)	(2)	(3)	(4)	(5)	(6)
% full-time employees	0.077 (0.047)	0.077 (0.047)	0.087* (0.043)	0.070 (0.045)	0.057 (0.057)	0.051 (0.056)
Log(# of employees)	0.075*** (0.009)	0.074*** (0.009)	0.064*** (0.007)	0.046*** (0.010)	0.068*** (0.011)	0.047*** (0.012)
Importance to recruiting of...						
References		-0.009 (0.033)	0.086** (0.029)	-0.019 (0.031)	-0.021 (0.040)	-0.020 (0.038)
Skills		0.038 (0.048)	-0.053 (0.034)	0.021 (0.045)	0.050 (0.057)	0.029 (0.053)
Referrals		-0.018 (0.030)	-0.027 (0.026)	-0.016 (0.028)	0.016 (0.037)	0.012 (0.034)
Qualifications or experience		0.020 (0.050)	0.009 (0.039)	-0.032 (0.047)	-0.032 (0.059)	-0.039 (0.054)
Motivation		0.038 (0.036)	0.096** (0.032)	0.021 (0.034)	-0.003 (0.046)	-0.010 (0.043)
Personality		0.036 (0.071)	0.069 (0.054)	0.024 (0.061)	0.009 (0.081)	0.016 (0.070)
Performance or competence test used				0.231*** (0.028)		0.229*** (0.034)
Recruits using internet					0.088** (0.033)	0.052 (0.033)
Recruits using print ads					-0.068* (0.034)	-0.083** (0.032)
# of firms	2,622	2,622	2,288	2,622	2,129	2,129

Notes: The table reports estimated marginal effects from Probit models of the probability of using personality tests in hiring. The dependent variable is an indicator for the use of personality or attitude tests when filling vacancies. The standard errors are estimated using the Taylor-linearized variance estimation method as recommended in the WERS documentation. Columns (1) and (2) and (4) to (6) use the 2011 WERS sample; Column (3) uses the 2004 WERS sample. The estimation sample in Columns (5) and (6) is restricted to employers who had hired in their largest occupational group in the last year. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Appendix A: Instructions, Job Descriptions, and Scales

Instructions (Baseline)

Please answer the following questions about yourself. Pay careful attention to response scales provided at the top of the screen as they may change from question to question.

You will be paid **\$7** for participating in this experiment. Are there any questions?

Your participation in today's experiment makes you eligible for another experiment to be held in the future. You will receive an invitation to this experiment.

Instructions (Evaluation: Extroversion, Introversion, No Priming)

This is an experiment in the economics of decision making.

Each subject will be randomly assigned to a group of 3-5 subjects (most groups will have 4 subjects) ; each group will be assigned a job description. All members of a group will receive the same job description. We will then administer an intelligence test and a personality test to determine who to "hire" for the job. We will weigh these two tests according to some criteria and one subject within each group who best meets these criteria will receive a bonus of \$25; the remaining subjects in each group will not receive any bonus.

After the bonus has been awarded, you will be asked to answer some further questions. You will be paid for one portion of the additional questions, which will be explained to you at the time.

Your earnings today will have four components.

- 1) You will be paid \$0.20 for every correct answer on the intelligence test.
- 2) The subject who is "hired" from his or her group will receive a bonus payment of \$25.
- 3) You will be paid for one portion of the additional questions.
- 4) All subjects who participate will receive a \$7 show-up fee.

Are there any questions?

We give you a couple minutes now to read the job description before we begin the intelligence test. You will then have 20 minutes to complete the intelligence test.

Instructions (Evaluation: Audit)

This is an experiment in the economics of decision making.

Each subject will be randomly assigned to a group of 3-5 subjects (most groups will have 4 subjects) ; each group will be assigned a job description. All members of a group will receive the same job description. We will then administer an intelligence test and a personality test to determine who to "hire" for the job. We will weigh these two tests according to some criteria. One subject within each group who best meets the criteria will be eligible to receive a bonus of \$25; the remaining subjects in each group will not receive any bonus.

All subjects' personality test answers will be subject to a 50% probability of being audited. This means that if you were to take the personality test many times, you could expect to be audited half of the time. Whether or not you are audited is determined randomly. Being audited means that we will compare your responses to the personality test to those you provided in an earlier session to determine your eligibility for the bonus. The audit procedure is described below.

The subject who best meets the criteria for the job is "hired" will receive the bonus if either a) he/she is not audited, or b) he/she is audited and "passes" the audit. If the "hired" subject fails the audit, the "hired" subject will not receive the \$25 bonus.

After the audits occur, you will be asked to answer some further questions. You will be paid for one portion of the additional questions, which will be explained to you at the time.

Audits

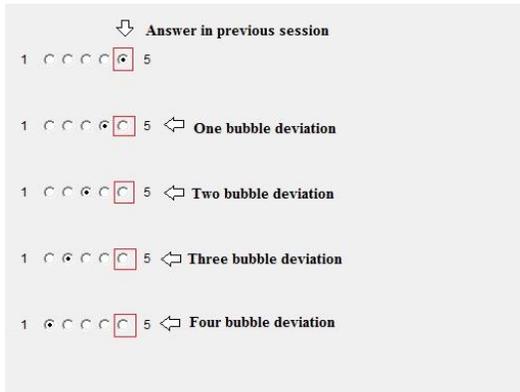
All audits will occur *after* we have determined who to "hire" for the job based on the intelligence and personality tests. If you are randomly selected to be audited, we will compare your answers from the previous session in which you completed the personality test to your answers today for 8 randomly selected questions. If your answers to those 8 questions differ from your previous answers by a total of 8 "bubble deviations" or more, you fail the audit.

If you are selected to be “hired” and you fail an audit, you will not receive the bonus. Failing the audit has NO effect on your earnings if you are not selected to be “hired.” Recall that on average about half of subjects will be audited.

A bubble deviation is the number of bubbles between the bubble you filled in previously and the bubble you fill in today, regardless of whether your answer for today is to the right or the left of your answer in the previous session.

Below are some example of bubble deviations:

Example 1



↓ Answer in previous session

1 ○ ○ ○ ○ 4 5

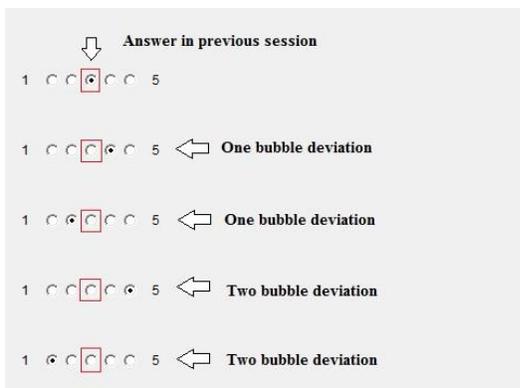
1 ○ ○ ○ ○ 3 5 ⇔ One bubble deviation

1 ○ ○ ○ ○ 2 5 ⇔ Two bubble deviation

1 ○ ○ ○ ○ 1 5 ⇔ Three bubble deviation

1 ○ ○ ○ ○ 5 5 ⇔ Four bubble deviation

Example 2



↓ Answer in previous session

1 ○ ○ 3 ○ ○ 5

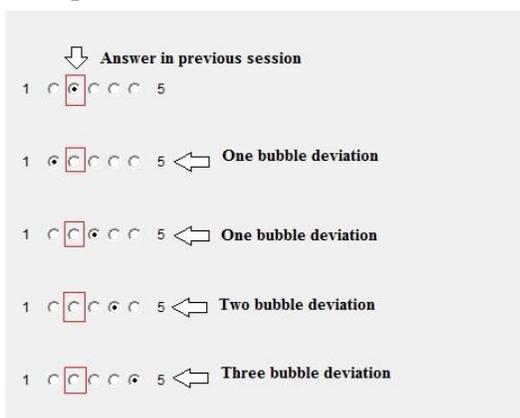
1 ○ ○ 4 ○ ○ 5 ⇔ One bubble deviation

1 ○ ○ 2 ○ ○ 5 ⇔ One bubble deviation

1 ○ ○ 1 ○ ○ 5 ⇔ Two bubble deviation

1 ○ ○ 5 ○ ○ 5 ⇔ Two bubble deviation

Example 3



↓ Answer in previous session

1 ○ 2 ○ ○ ○ ○ 5

1 ○ 1 ○ ○ ○ ○ 5 ⇔ One bubble deviation

1 ○ 3 ○ ○ ○ ○ 5 ⇔ One bubble deviation

1 ○ 4 ○ ○ ○ ○ 5 ⇔ Two bubble deviation

1 ○ 5 ○ ○ ○ ○ 5 ⇔ Three bubble deviation

If your answers to the 8 randomly selected questions differ from your previous answers by a total of 8 “bubble deviations” or more, you fail the audit.

Your earnings today will have four components.

- 1) You will be paid \$0.20 for every correct answer on the intelligence test.
- 2) The subject who is “hired” from his or her group may receive a bonus payment of \$25.

- 3) You will be paid for one portion of the additional questions.
- 4) All subjects who participate will receive a \$7 show-up fee.

Are there any questions?

We give you a couple minutes now to read the job description before we begin the intelligence test. You will then have 20 minutes to complete the intelligence test.

Job descriptions (Extroversion and Audit)

Client Services Representative

About the Job

EXCELLENT OPPORTUNITY for a client services representative at a collegial and exciting staffing firm. This is an ideal position for an outgoing, friendly, energetic person who can represent our company to the outside world and help us stand out from the crowd.

Primary Duties:

Accomplish marketing and sales objectives by taking charge of programs to expand our customer base
Build relationships with key accounts by making regular visits
Maintain personal networks, participate in networking and professional organizations.
Implementation of business-need assessment programs for clients
Provide staffing support to clients to help them meet their business needs
Allocate firm resources across a diverse set of clients

An ideal candidate will be someone who:

Is a happy, cheerful, optimistic, and enthusiastic “go-getter”
Is proficient at building and maintaining relationships with client firms and staffers
Enjoys interacting with customers and the public
Is a fun-to-be-around person who customers will want to invite to backyard barbecues.
Thrives in crowds.
Has boundless energy and vigor to enthusiastically promote our firm at every opportunity.
Is aggressive and assertive in ensuring that our clients’ problems get resolved in a timely fashion
Speaks up and takes charge to resolve problems
Is happy about working at our firm!
Is able to take initiative, be assertive, and follow projects to completion
Has great people skills
Is proficient in Microsoft Office (Word, Excel, PowerPoint, etc.)
Has excellent project management skills
Has strong analytical skills
Has strong communication skills
Has financial acumen

Skills/Qualifications:

Bachelor’s Degree
Two years of experience in a related field
Valid driver’s license upon starting work

Job descriptions (Introversion)

Client Services Representative

About the Job

EXCELLENT OPPORTUNITY for a client services representative at a low-key staffing firm. This is an ideal position for a contemplative, reserved, independent person who can represent our company and grow.

Primary Duties:

Accomplish marketing and sales objectives by developing programs that our customers implement remotely
Assist key accounts while working from home, often at off-peak times
Keeping up-to-date on industry developments without regular contact from supervisors or co-workers
Implementation of business-need assessment programs for clients
Provide staffing support to clients to help them meet their business needs
Allocate firm resources across a diverse set of clients

An ideal candidate will be someone who:

Enjoys tackling projects solo
Can get the job done without making a splash or interrupting clients normal business
Is capable of devising and revising business strategies independently
Gets a kick out of solving puzzles
Is proficient in Microsoft Office (Word, Excel, PowerPoint, etc.)
Has excellent project management skills
Has strong analytical skills
Has strong communication skills
Has financial acumen

Skills/Qualifications:

Bachelor's Degree
Two years of experience in a related field
Valid driver's license upon starting work

Job descriptions (No Priming)

Client Services Representative

About the Job

EXCELLENT OPPORTUNITY for a client services representative at a staffing firm.

Primary Duties:

Implementation of business-need assessment programs for clients
Provide staffing support to clients to help them meet their business needs
Allocate firm resources across a diverse set of clients

An ideal candidate will be someone who has:

Is proficient in Microsoft Office (Word, Excel, PowerPoint, etc.)
Has excellent project management skills
Has strong analytical skills
Has financial acumen

Skills/Qualifications:

Bachelor's Degree
Two years of experience in a related field
Valid driver's license upon starting work

Scales

Big 5

Neuroticism

+ keyed

Get angry easily.
Get upset easily.
Change my mood a lot.
Am a person whose moods go up and down easily.
Get easily agitated.

– keyed

Can be stirred up easily.
 Am filled with doubts about things.
 Feel threatened easily.
 Worry about things.
 Am easily discouraged.
 Become overwhelmed by events.
 Am afraid of many things.

Seldom feel blue.
 Feel comfortable with myself.
 Rarely feel depressed.
 Am not embarrassed easily.
 Rarely get irritated.
 Keep my emotions under control.
 Rarely lose my composure.
 Am not easily annoyed.

Agreeableness

+ keyed

Feel others' emotions.
 Inquire about others' well-being.
 Sympathize with others' feelings.
 Take an interest in other people's lives.
 Like to do things for others.
 Respect authority.
 Hate to seem pushy.
 Avoid imposing my will on others.
 Rarely put people under pressure.

– keyed

Insult people.
 Believe that I am better than others.
 Take advantage of others.
 Seek conflict.
 Love a good fight.
 Am out for my own personal gain.
 Am not interested in other people's problems.
 Can't be bothered with other's needs.
 Am indifferent to the feelings of others.
 Take no time for others.
 Don't have a soft side.

Conscientiousness

+ keyed

Carry out my plans.
 Finish what I start.
 Get things done quickly.
 Always know what I am doing.
 Like order.
 Keep things tidy.
 Follow a schedule.
 Want everything to be "just right."
 See that rules are observed.
 Want every detail taken care of.

– keyed

Leave my belongings around.
 Am not bothered by messy people.
 Am not bothered by disorder.
 Dislike routine.
 Waste my time.
 Find it difficult to get down to work.
 Mess things up.
 Don't put my mind on the task at hand.
 Postpone decisions.
 Am easily distracted.

Extraversion

+ keyed

Make friends easily.
 Warm up quickly to others.
 Show my feelings when I'm happy.
 Have a lot of fun.
 Laugh a lot.
 Take charge.
 Have a strong personality.
 Know how to captivate people.
 See myself as a good leader.
 Can talk others into doing things.
 Am the first to act.

– keyed

Do not have an assertive personality.
 Lack the talent for influencing people.
 Wait for others to lead the way.
 Hold back my opinions.
 Am hard to get to know.
 Keep others at a distance.
 Reveal little about myself.
 Rarely get caught up in the excitement.
 Am not a very enthusiastic person.

Openness/Intellect

+ keyed	Am quick to understand things. Can handle a lot of information. Like to solve complex problems. Have a rich vocabulary. Think quickly. Formulate ideas clearly.
+ keyed	Enjoy the beauty of nature. Believe in the importance of art. Love to reflect on things. Get deeply immersed in music. See beauty in things that others might not notice. Need a creative outlet.
- keyed	Do not like poetry. Seldom get lost in thought. Seldom daydream. Seldom notice the emotional aspects of paintings and pictures. Have difficulty understanding abstract ideas. Avoid philosophical discussions. Avoid difficult reading material. Learn things slowly.

Scoring instructions:

For + keyed items, the response "Very Inaccurate" is assigned a value of 1, "Moderately Inaccurate" a value of 2, "Neither Inaccurate nor Accurate" a 3, "Moderately Accurate" a 4, and "Very Accurate" a value of 5.

For - keyed items, the response "Very Inaccurate" is assigned a value of 5, "Moderately Inaccurate" a value of 4, "Neither Inaccurate nor Accurate" a 3, "Moderately Accurate" a 2, and "Very Accurate" a value of 1.

Once numbers are assigned for all of the items in the scale, just sum all the values to obtain a total scale score.

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

Machiavellianism

+ keyed	Find it easy to manipulate others. Have a natural talent for influencing people. Can talk others into doing things.
- keyed	Find it difficult to manipulate others. Hate being the center of attention. Lack the talent for influencing people.

Scoring instructions:

For + keyed items, the response "Very Inaccurate" is assigned a value of 1, "Moderately Inaccurate" a value of 2, "Neither Inaccurate nor Accurate" a 3, "Moderately Accurate" a 4, and "Very Accurate" a value of 5.

For - keyed items, the response "Very Inaccurate" is assigned a value of 5, "Moderately Inaccurate" a value of 4, "Neither Inaccurate nor Accurate" a 3, "Moderately Accurate" a 2, and "Very Accurate" a value of 1.

Once numbers are assigned for all of the items in the scale, just sum all the values to obtain a total scale score.

Jackson, D. N. (1994). *Jackson Personality Inventory-Revised manual*. Port Huron, MI: Sigma Assessment Systems

Self-monitoring

+ keyed	Would make a good actor. Put on a show to impress people. Am likely to show off if I get the chance. Am the life of the party. Am good at making impromptu speeches. Like to attract attention. Use flattery to get ahead.
- keyed	Hate being the center of attention. Would not be a good comedian. Don't like to draw attention to myself.

Scoring instructions:

For + keyed items, the response "Very Inaccurate" is assigned a value of 1, "Moderately Inaccurate" a value of 2, "Neither Inaccurate nor Accurate" a 3, "Moderately Accurate" a 4, and "Very Accurate" a value of 5.

For - keyed items, the response "Very Inaccurate" is assigned a value of 5, "Moderately Inaccurate" a value of 4, "Neither Inaccurate nor Accurate" a 3, "Moderately Accurate" a 2, and "Very Accurate" a value of 1.

Once numbers are assigned for all of the items in the scale, just sum all the values to obtain a total scale score.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology, 30*, 526-537

Behavioral Inventory of Desirable Responding

My first impressions of people usually turn out to be right.
It would be hard for me to break any of my bad habits.
I don't care to know what other people really think of me.
I have not always been honest with myself.
I always know why I like things.
When my emotions are aroused, it biases my thinking.
Once I've made up my mind, other people can seldom change my opinion.
I am not a safe driver when I exceed the speed limit.
I am fully in control of my own fate.
It's hard for me to shut off a disturbing thought.
I never regret my decisions.
I sometimes lose out on things because I can't make up my mind soon enough.
The reason I vote is because my vote can make a difference.
My parents were not always fair when they punished me.
I am a completely rational person.
I rarely appreciate criticism.
I am very confident of my judgments.
I have sometimes doubted my ability as a lover.
It's all right with me if some people happen to dislike me.
I don't always know the reasons why I do the things I do.
I sometimes tell lies if I have to.
I never cover up my mistakes.
There have been occasions when I have taken advantage of someone.
I never swear.
I sometimes try to get even rather than forgive and forget.
I always obey laws, even if I'm unlikely to get caught.
I have said something bad about a friend behind his/her back.
When I hear people talking privately, I avoid listening.
I have received too much change from a salesperson without telling him or her.
I always declare everything at customs.
When I was young I sometimes stole things.
I have never dropped litter on the street.
I sometimes drive faster than the speed limit.
I never read sexy books or magazines.
I have done things that I don't tell other people about.
I never take things that don't belong to me.
I have taken sick-leave from work or school even though I wasn't really sick.
I have never damaged a library book or store merchandise without reporting it.
I have some pretty awful habits.
I don't gossip about other people's business.

Self Deceptive Enhancement (SDE): Items 1 – 20 (Reverse scored items: 2,4,6,8,10,12,14,16,18,20).

Impression Management (IM): Items 21 – 40 (Reverse scored items: 21,23,25,27,29,31,33,35,37,39).

Scoring instructions:

Each + keyed item is scored on a 7-point Likert scale with "Not true" assigned a value of 1, "Somewhat true" assigned a value of 4, and "Very true" assigned a value of 7.

For - keyed items, the Likert ratings are reversed.

For each subscale, add one point for every 6 or 7, then sum the number of points

Paulhus, D. L. (1991). *Balanced Inventory of Desirable Responding (BIDR) reference manual for version 6. Manual available from author at Department of Psychology, University of British Columbia, Vancouver, BC, Canada V6T 1Y7.*

Abbreviated 4-item Rotter Internal-External Locus of Control Scale

- A. What happens to me is my own doing.
B. Sometimes I feel that I don't have enough control over the direction my life is taking.
- A. When I make plans, I am almost certain that I can make them work.
B. It is not always wise to plan too far ahead because many things turn out to be a matter of good or bad fortune.
- A. In my case getting what I want has little or nothing to do with luck.
B. Many times we might just as well decide what to do by flipping a coin.
- A. Many times I feel that I have little influence over the things that happen to me.
B. It is impossible for me to believe that chance or luck plays an important role in my life.

Scoring instructions:

Respondents choose which statement, either A or B, is closer to their opinion for each pair of statements. Respondents then choose whether the statement is "much closer" to their opinion or "slightly closer."

Choosing the external (fatalistic) statement and stating that it is "much closer" to their opinion is given a 1, choosing the external (fatalistic) statement and stating that it is "slightly closer" to their opinion is given a 2, choosing the internal (efficacious) statement and stating that it is "slightly closer" to their opinion is given a 3, and choosing the internal (efficacious) statement and stating that it is "much closer" to their opinion is given a 4. Total scores are the sum of the scores assigned to each pair of statements.

Optimism (Lot-R)

1. In uncertain times, I usually expect the best.
2. If something can go wrong for me, it will.
3. I'm always optimistic about my future.
4. I hardly ever expect things to go my way.
5. I rarely count on good things happening to me.
6. Overall, I expect more good things to happen to me than bad.

Scoring instructions:

The response "I DISagree a lot" is assigned a value of 1, "I DISagree a little" a value of 2, "I neither agree nor disagree" a 3, "I agree a little" a 4, and "I agree a lot" a value of 5.

Once numbers are assigned for all of the items in the scale, just sum all the values to obtain a total scale score.

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A re-evaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, 67, 1063-1078

Holt-Laury (Low Stakes)

In the questions that follow, you are going to be asked to make ten decisions. Each decision will be between Option A and Option B. Please enter your decisions below and on the corresponding sheet that was handed out to you. Only one of the ten choices you make will be used to determine your earnings for this part of the experiment. After you answer all 10 questions you will be shown the "decision selected" and "outcome" which will be used to calculate your earnings. Be sure to write these down. Each decision is a paired choice between "Option A" and "Option B." You will make ten choices. Before you start making your ten choices, let me explain what these choices mean. Imagine a ten-sided die that will be used to determine payoffs; the faces are numbered from 1 to 10. After you have made all of your choices, the die would be thrown twice, once to select one of the ten decisions to be used, and a second time to determine what your payoff is for the option you chose, A or B, for the particular decision selected. Given this, you should make the choice that you would prefer if we were throwing the die for real. Now, please look at Decision 1 at the top. Option A pays 200 pennies if the throw of the ten sided die is 1, and it pays 160 pennies if the throw is 2-10. Option B yields 385 pennies if the throw of the die is 1, and it pays 10 pennies if the throw is 2-10. The other Decisions are similar, except that as you move down the table, the chances of the higher payoff for each option increase. In fact, for Decision 10 in the bottom row, the die will not be needed since each option pays the highest payoff for sure, so your choice here is between 200 pennies or 385 pennies.

To summarize, you will make ten choices: for each decision row you will have to choose between Option A and Option B. You may choose A for some decision rows and B for other rows, and you may change your decisions and make them in any order.

	Option A	Option B	Your Choice
1.	1/10 of \$2.00	9/10 of \$1.60	1/10 of \$3.85 9/10 of \$0.10 A / B
2.	2/10 of \$2.00	8/10 of \$1.60	2/10 of \$3.85 8/10 of \$0.10 A / B
3.	3/10 of \$2.00	7/10 of \$1.60	3/10 of \$3.85 7/10 of \$0.10 A / B
4.	4/10 of \$2.00	6/10 of \$1.60	4/10 of \$3.85 6/10 of \$0.10 A / B
5.	5/10 of \$2.00	5/10 of \$1.60	5/10 of \$3.85 5/10 of \$0.10 A / B
6.	6/10 of \$2.00	4/10 of \$1.60	6/10 of \$3.85 4/10 of \$0.10 A / B
7.	7/10 of \$2.00	3/10 of \$1.60	7/10 of \$3.85 3/10 of \$0.10 A / B
8.	8/10 of \$2.00	2/10 of \$1.60	8/10 of \$3.85 2/10 of \$0.10 A / B
9.	9/10 of \$2.00	1/10 of \$1.60	9/10 of \$3.85 1/10 of \$0.10 A / B
10.	10/10 of \$2.00	0/10 of \$1.60	10/10 of \$3.85 0/10 of \$0.10 A / B

Lying scenarios

A co-worker of Melinda is hosting a party and asks Melinda if she is enjoying the food. In order not to hurt his feelings, Melinda lies and says the food is fantastic, even though it is overcooked and tasteless.

Mike is working on a group project with another student who comes up with an idea for their project. In order to avoid conflict, Mike lies and says he likes the idea, even though he thinks the idea is a poor one.

On a visit to another country, Lea buys some gold jewelry. In order to avoid paying duty on the jewelry, Lea lies and tells the customs official that she did not buy anything while in the country.

Sean accidentally backs into a parked car. As he is driving away, the owner arrives and asks Sean if he saw who damaged his car. In order to avoid paying for the damage, Sean lies and says he has no idea who did it.

Jamie's friend really wants her to go to a concert next weekend. Jamie would rather spend the weekend on her own. In order to avoid causing a conflict with her friend, Jamie lies and says she has to work that weekend.

Michelle's co-worker is very upset about a new policy at work. In order to avoid an argument, Michelle lies and agrees that the policy is unreasonable, even though she strongly approves of the policy.

Harry is overburdened at work and has little time for his family. A new co-worker asks Harry if he has time to help him learn the new bookkeeping system. In order to help him out, Harry lies and says he has time, even though he doesn't.

Mary's new co-worker asks for advice on applying for a position that has opened up in the company. Because Mary hopes to get the position herself, she lies and says the position has already been promised to someone else.

Kate's fellow students are complaining about an instructor they don't like. In order to fit in, Kate lies and says she dislikes the instructor as well, even though she really likes the instructor.

Tom and a friend buy a lottery ticket together. When Tom takes the ticket in to check if they won anything, he receives \$200. In order to keep all the money for himself, Tom lies to his friend and tells him they won nothing.

Terry's new friend hates hunting. In order to be liked by her, Terry lies and tells her he has never hunted, even though he is an avid hunter.

Bob's neighbour asks if he will vote for him in the upcoming election. In order to avoid conflict, Bob lies and says he will, even though he intends to vote for another candidate.

One day Jerry is drinking with some co-workers who start talking about their experiences playing hockey. In order to fit in Jerry lies and tells stories about playing hockey himself, even though he has never actually played hockey.

Brad's friend asks if he will help her move the next day. In order to be helpful, Brad lies and tells her that he has nothing planned and will help her move, even though he had booked in to work that day.

Kira's friend tells her she really likes a new political party and asks Kira if she likes the party. In order to gain her friend's approval, she lies and says she does like them, even though she really dislikes the party.

Susan's friend at work asks Susan to write a letter of reference for her. To help her friend out, Susan lies in the letter and says she believes her friend is perfect for the job, even though Susan has some reservations about her friend's ability to do the job.

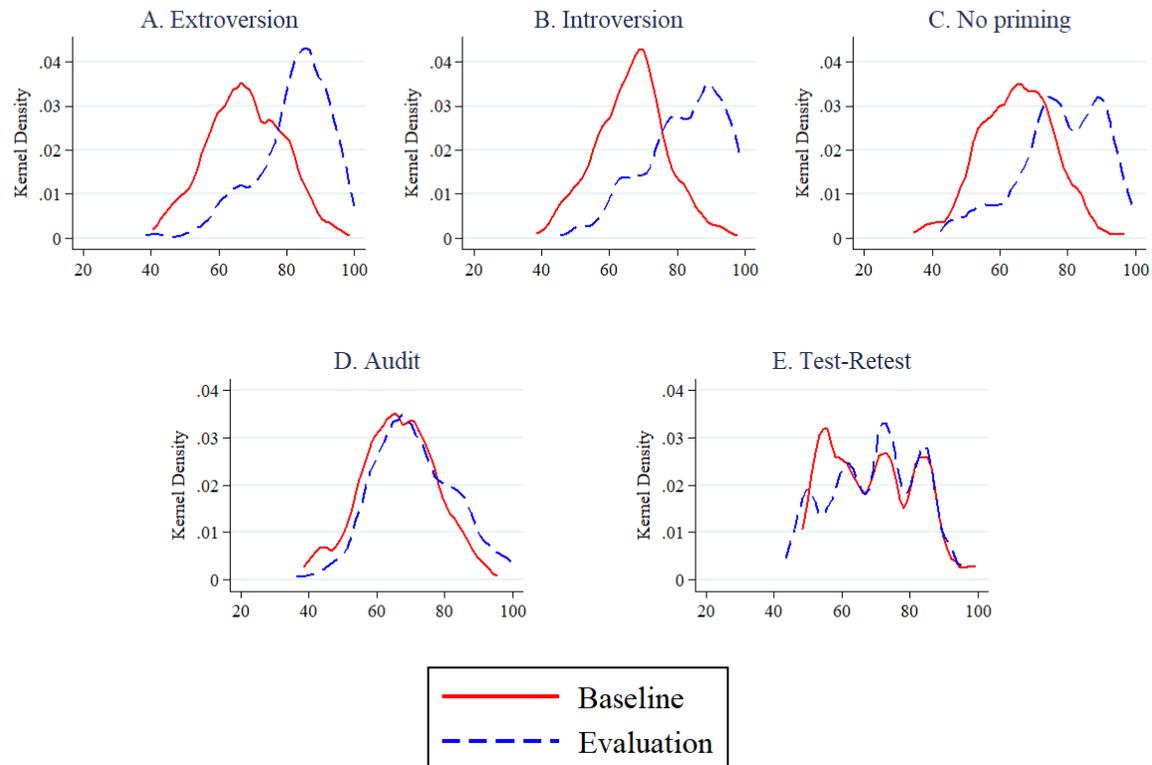
Scoring instructions:

Respondents rated how acceptable it was for the person in the scenario to have lied using nine-point Likert scales ranging from extremely unacceptable to extremely. For each respondent, one acceptability score for each type of lie (altruistic, conflict avoidance, social acceptance, and self-gain) was obtained by calculating his/her average score across the four relevant scenarios.

Scenarios 1, 7, 14, and 16 are altruistic lies, scenarios 2, 5, 6, and 12 are lies to avoid conflict, scenarios 9, 11, 13, and 15 are lies to gain social acceptance, and scenarios 3, 4, 8, and 10 are lies told to benefit the liar.

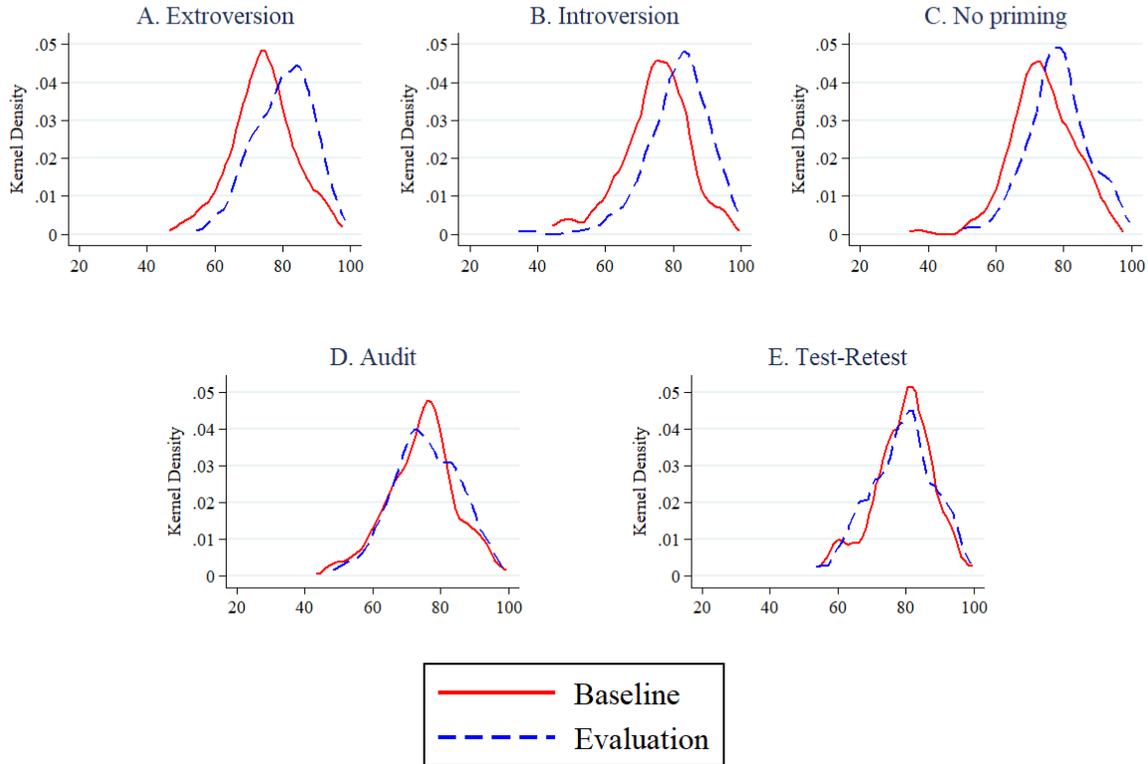
McLeod, B. A., & Genereux, R. L. (2008). Predicting the acceptability and likelihood of lying: The interaction of personality with type of lie. *Personality and individual differences, 45*(7), 591-596.

Appendix B Figure 1: Conscientiousness kernel densities



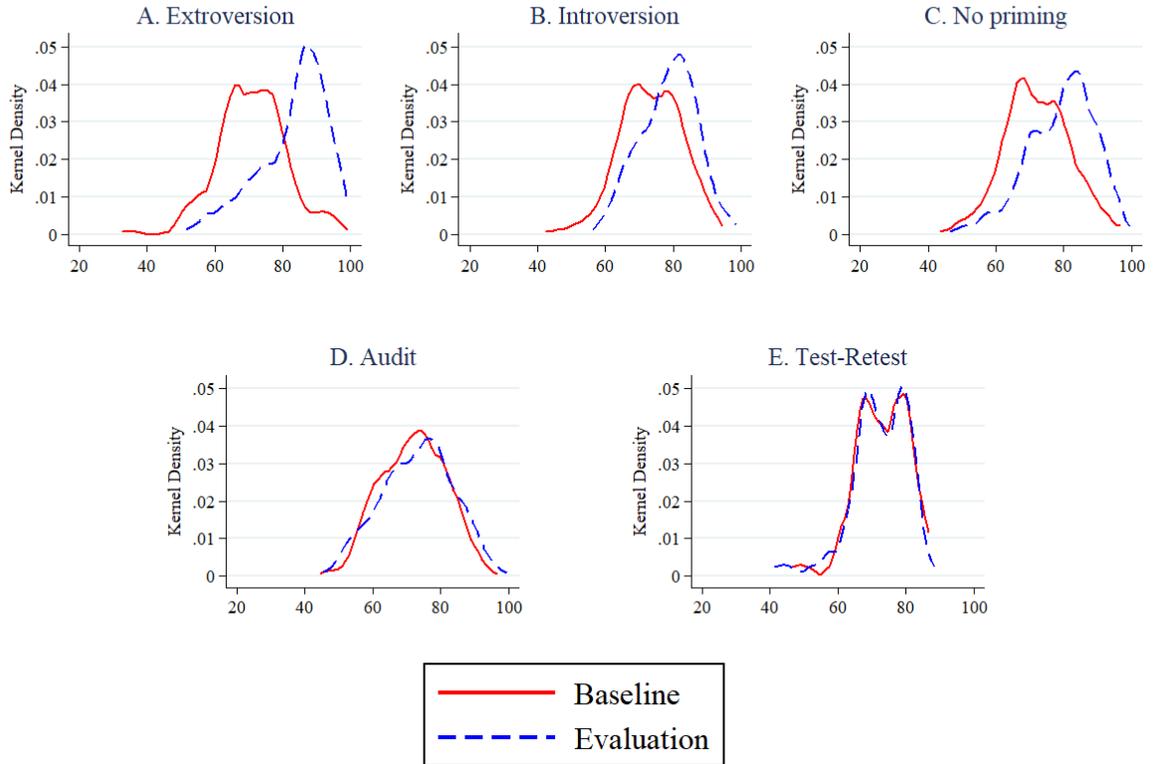
Notes: Kernel density plots of raw conscientiousness scores from the Baseline (non-incentivized) and Evaluation (incentivized) sessions by treatment are displayed. Panel A depicts the densities for the *Extroversion* treatment, Panel B the *Introversion* treatment, Panel C the *No Priming* treatment, Panel D the *Audit* treatment and Panel E the *Test-Retest* treatment. The Evaluation session in the *Test-Retest* treatment was the same as the Baseline session and thus not incentivized. The p-values for the two-sample Kolmogorov-Smirnov tests for the equality of the distributions between the Baseline and Evaluation sessions are 0.000, 0.000, 0.000, and 0.086 for Panels A, B, C, and D, respectively. The estimated probability that a random draw from the non-incentivized score distribution is larger than a random draw from the incentivized score distribution is 0.175 in Panel A, 0.167 in Panel B, 0.212 in Panel C, and 0.403 in Panel D.

Appendix B Figure 1: Agreeableness kernel densities



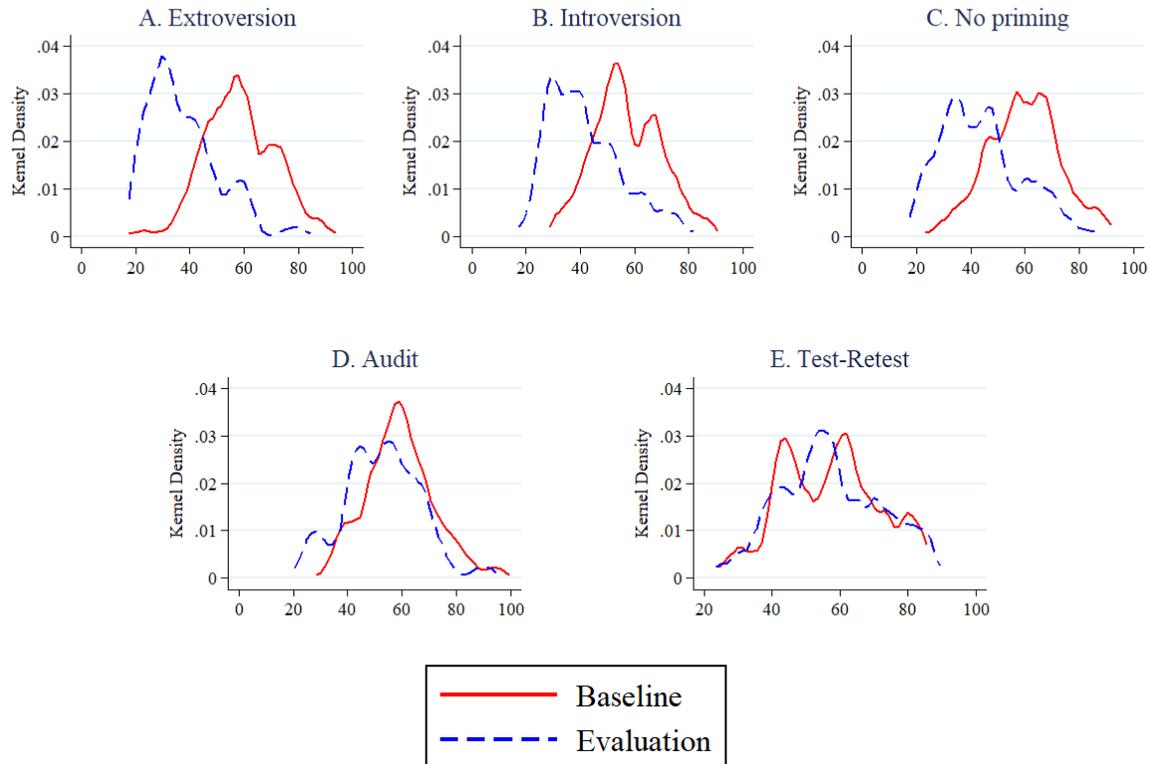
Notes: Kernel density plots of raw agreeableness scores from the Baseline (non-incentivized) and Evaluation (incentivized) sessions by treatment are displayed. Panel A depicts the densities for the *Extroversion* treatment, Panel B the *Introversion* treatment, Panel C the *No Priming* treatment, Panel D the *Audit* treatment and Panel E the *Test-Retest* treatment. The Evaluation session in the *Test-Retest* treatment was the same as the Baseline session and thus not incentivized. The p-values for the two-sample Kolmogorov-Smirnov tests for the equality of the distributions between the Baseline and Evaluation sessions are 0.000, 0.000, 0.000, and 0.357 for Panels A, B, C, and D, respectively. The estimated probability that a random draw from the non-incentivized score distribution is larger than a random draw from the incentivized score distribution is 0.313 in Panel A, 0.299 in Panel B, 0.361 in Panel C, and 0.466 in Panel D.

Appendix B Figure 1: Openness kernel densities



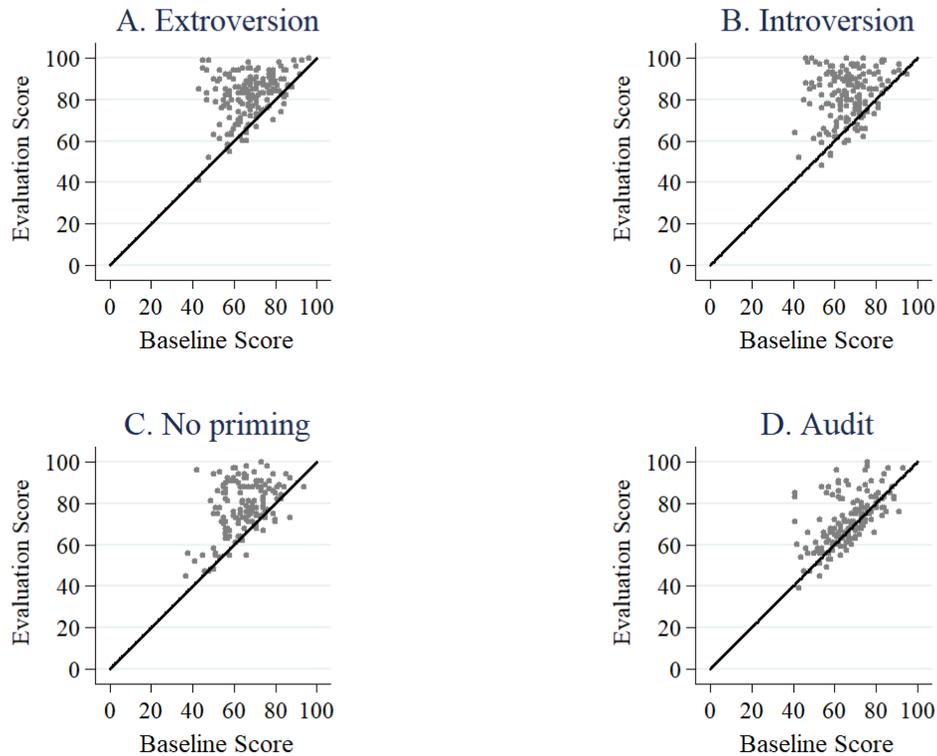
Notes: Kernel density plots of raw openness scores from the Baseline (non-incentivized) and Evaluation (incentivized) sessions by treatment are displayed. Panel A depicts the densities for the *Extroversion* treatment, Panel B the *Introversion* treatment, Panel C the *No Priming* treatment, Panel D the *Audit* treatment and Panel E the *Test-Retest* treatment. The Evaluation session in the *Test-Retest* treatment was the same as the Baseline session and thus not incentivized. The p-values for the two-sample Kolmogorov-Smirnov tests for the equality of the distributions between the Baseline and Evaluation sessions are 0.000, 0.000, 0.000, and 0.606 for Panels A, B, C, and D, respectively. The estimated probability that a random draw from the non-incentivized score distribution is larger than a random draw from the incentivized score distribution is 0.194 in Panel A, 0.338 in Panel B, 0.290 in Panel C, and 0.468 in Panel D.

Appendix B Figure 1: Neuroticism kernel densities



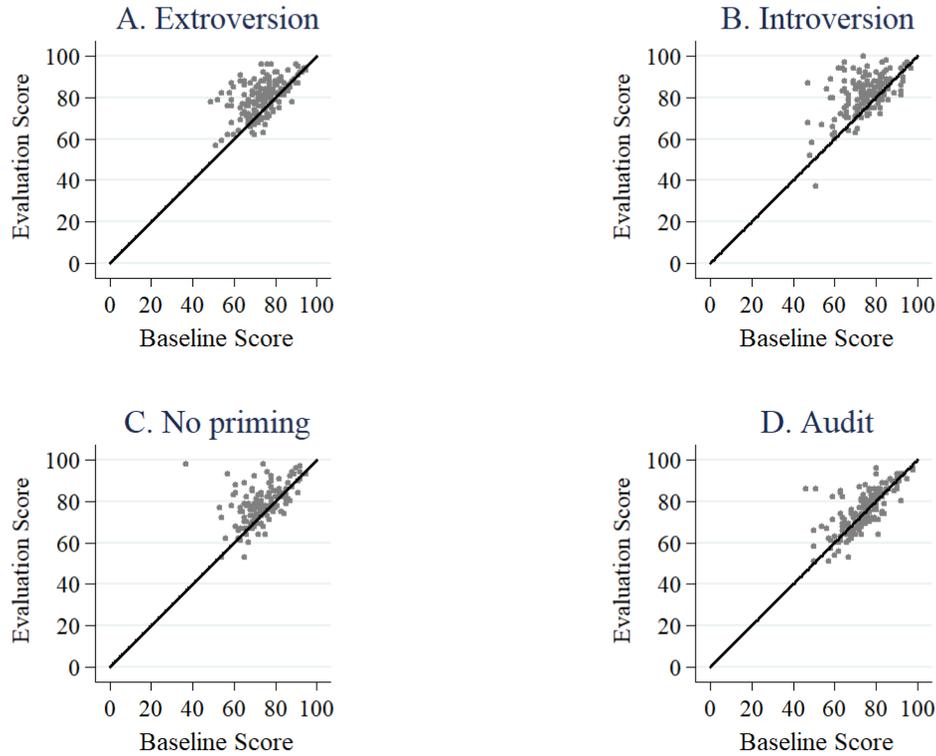
Notes: Kernel density plots of raw neuroticism scores from the Baseline (non-incentivized) and Evaluation (incentivized) sessions by treatment are displayed. Panel A depicts the densities for the *Extroversion* treatment, Panel B the *Introversion* treatment, Panel C the *No Priming* treatment, Panel D the *Audit* treatment and Panel E the *Test-Retest* treatment. The Evaluation session in the *Test-Retest* treatment was the same as the Baseline session and thus not incentivized. The p-values for the two-sample Kolmogorov-Smirnov tests for the equality of the distributions between the Baseline and Evaluation sessions are 0.000, 0.000, 0.000, and 0.001 for Panels A, B, C, and D, respectively. The estimated probability that a random draw from the non-incentivized score distribution is larger than a random draw from the incentivized score distribution is 0.876 in Panel A, 0.807 in Panel B, 0.784 in Panel C, and 0.632 in Panel D.

Appendix B Figure 2: Conscientiousness scatter plots



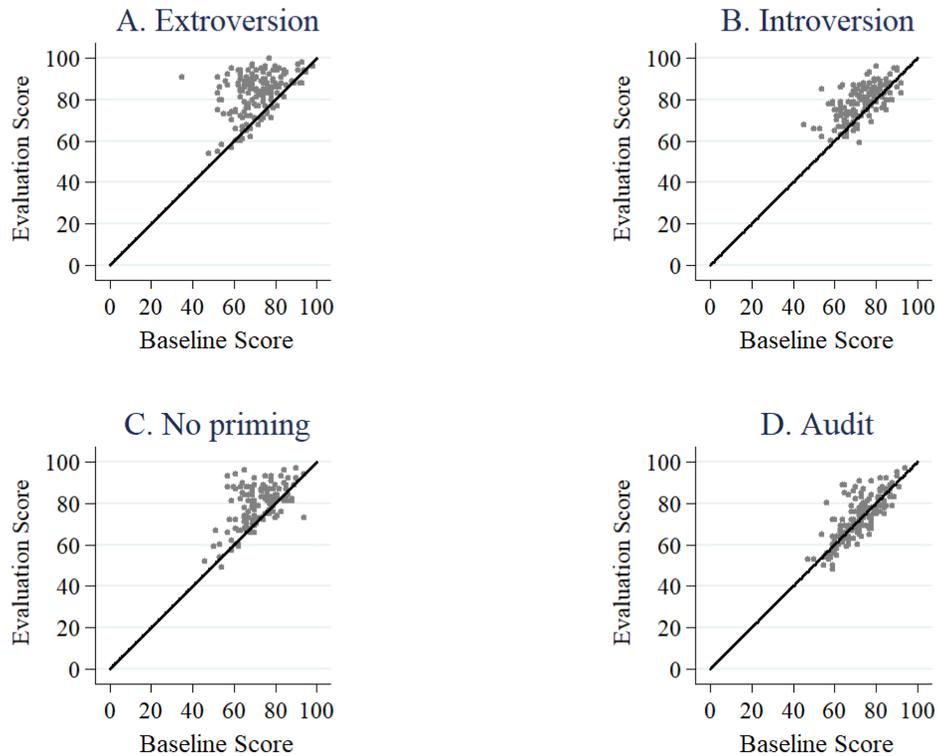
Notes: The figure displays scatterplots of the raw, Baseline session (non-incentivized) conscientiousness score on the x-axis against the raw, Evaluation session (incentivized) conscientiousness score on the y-axis by treatment. The plots in each panel also depict a 45-degree line for use as a reference. Panel A depicts the scatterplot for the *Extroversion* treatment, Panel B the *Introversion* treatment, Panel C the *No Priming* treatment, and Panel D the *Audit* treatment. The Evaluation session in the *Test-Retest* treatment was the same as the Baseline session and thus not incentivized.

Appendix B Figure 2: Agreeableness scatter plots



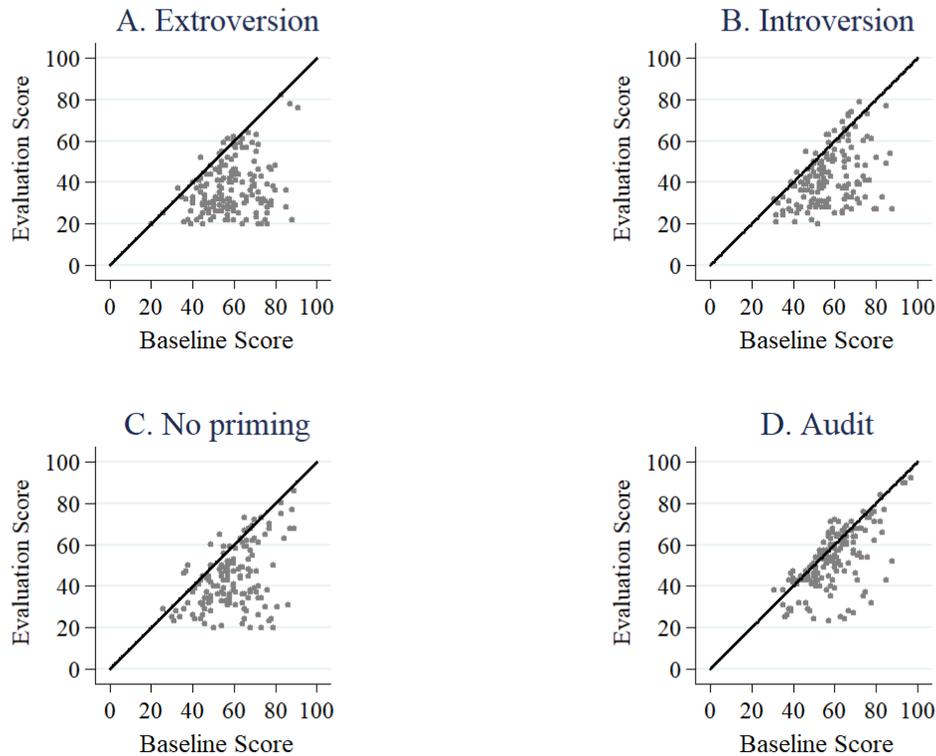
Notes: The figure displays scatterplots of the raw, Baseline session (non-incentivized) agreeableness score on the x-axis against the raw, Evaluation session (incentivized) agreeableness score on the y-axis by treatment. The plots in each panel also depict a 45-degree line for use as a reference. Panel A depicts the scatterplot for the *Extroversion* treatment, Panel B the *Introversion* treatment, Panel C the *No Priming* treatment, and Panel D the *Audit* treatment. The Evaluation session in the *Test-Retest* treatment was the same as the Baseline session and thus not incentivized.

Appendix B Figure 2: Openness scatter plots



Notes: The figure displays scatterplots of the raw, Baseline session (non-incentivized) openness score on the x-axis against the raw, Evaluation session (incentivized) openness score on the y-axis by treatment. The plots in each panel also depict a 45-degree line for use as a reference. Panel A depicts the scatterplot for the *Extroversion* treatment, Panel B the *Introversion* treatment, Panel C the *No Priming* treatment, and Panel D the *Audit*. The Evaluation session in the *Test-Retest* treatment was the same as the Baseline session and thus not incentivized.

Appendix B Figure 2: Neuroticism scatter plots



Notes: The figure displays scatterplots of the raw, Baseline session (non-incentivized) neuroticism score on the x-axis against the raw, Evaluation session (incentivized) neuroticism score on the y-axis by treatment. The plots in each panel also depict a 45-degree line for use as a reference. Panel A depicts the scatterplot for the *Extroversion* treatment, Panel B the *Introversion* treatment, Panel C the *No Priming* treatment, and Panel D the *Audit* treatment. The Evaluation session in the *Test-Retest* treatment was the same as the Baseline session and thus not incentivized.

Appendix B Table 1: Comparing subjects who return for the Evaluation session to subjects who do not return

	Non-Returnees	Returnees	Wilcoxon rank-sum p-value	Kolmogorov-Smirnov p-value
	(1)	(2)	(3)	(4)
Male	0.49	0.48	0.885	0.999
Age	20.75 (2.51)	21.04 (3.69)	0.708	0.999
White	0.21	0.23	0.547	0.999
Self-deception	5.33 (3.27)	5.82 (3.42)	0.091	0.345
Impression management	5.70 (3.30)	6.01 (3.64)	0.514	0.565
Self-monitoring	29.14 (7.88)	28.23 (8.35)	0.223	0.282
Machiavellianism	18.97 (4.43)	18.65 (4.71)	0.389	0.637
Acceptability of lies for personal gain	3.53 (1.34)	3.49 (1.47)	0.775	0.286
Acceptability of lies to be kind	5.03 (1.35)	5.01 (1.39)	0.975	0.461
Acceptability of lies to avoid conflict	4.80 (1.43)	4.72 (1.42)	0.748	0.886
Acceptability of lies to gain social acceptance	4.17 (1.61)	4.03 (1.56)	0.314	0.341
Optimism	21.30 (4.51)	20.76 (4.72)	0.159	0.190
Extroversion	69.94 (10.19)	68.08 (11.34)	0.098	0.242
Conscientiousness	65.67 (11.31)	66.46 (10.82)	0.447	0.582
Agreeableness	75.01 (9.79)	74.28 (9.49)	0.530	0.299
Openness	72.74 (9.43)	72.06 (9.49)	0.386	0.843
Neuroticism	59.87 (13.21)	58.50 (12.71)	0.209	0.350
# of subjects	162	642		

Notes: Columns (1) and (2) report the means and standard deviations in parentheses for subjects who did not return for an evaluation session and subjects who returned, respectively. Column (3) reports p-values for Wilcoxon rank-sum test for the equality of the means in Columns (1) and (2), while Column (4) reports the p-values for the Kolmogorov-Smirnov test for equality of distribution functions.

Appendix B Table 2a: Incentivized conscientiousness regressions

	Treatment				p-value of (1)=(2)=(3)=0 (5)
	Extroversion	Introversion	No Priming	Audit	
	(1)	(2)	(3)	(4)	(5)
Non-incentivized conscientiousness	0.365*** (0.113)	0.183* (0.109)	0.396*** (0.104)	0.525*** (0.097)	0.000
IQ	0.074 (0.077)	-0.042 (0.091)	0.004 (0.077)	-0.121* (0.072)	0.765
Risk aversion	0.101 (0.089)	-0.025 (0.088)	0.026 (0.073)	-0.004 (0.066)	0.685
Locus of control	-0.001 (0.084)	0.254*** (0.074)	0.177** (0.080)	-0.022 (0.067)	0.001
Acceptability of lies for gain	0.042 (0.096)	0.064 (0.075)	-0.053 (0.094)	0.076 (0.087)	0.741
Self-deception	0.021 (0.102)	0.018 (0.113)	0.011 (0.085)	0.123 (0.091)	0.994
Impression management	-0.042 (0.098)	-0.036 (0.120)	-0.050 (0.095)	0.113 (0.077)	0.908
Self-monitoring	0.025 (0.105)	-0.073 (0.111)	0.003 (0.102)	0.094 (0.093)	0.920
Machiavellianism	0.054 (0.099)	0.189 (0.124)	0.011 (0.108)	0.016 (0.101)	0.454
Optimism	-0.039 (0.099)	-0.124 (0.090)	0.010 (0.093)	-0.144 (0.091)	0.562
R ²	0.152	0.139	0.223	0.392	

Notes: Columns (1) to (4) report coefficient estimates from separate OLS regressions for each treatment. The dependent variable in each regression is the incentivized conscientiousness score from the Evaluation session standardized using the mean and standard deviation of the incentivized scores from subjects in the same treatment. The non-incentivized conscientiousness score used in the controls is standardized using the mean and standard deviation for the non-incentivized scores from subjects in the same treatment. The remaining controls—all of which were measured in the Baseline session except for locus of control—are standardized using the full sample of all subjects in all treatments. The standard errors reported in parentheses are scaled by the square of the observation's variance estimate (i.e., the diagonal element of the projection matrix) as suggested by Davidson and MacKinnon [1993] and implemented in Stata using the “vce(hc3)” option. Column (5) reports the p-value for a heteroscedasticity-robust Wald test for the hypotheses that the coefficients in Columns (1) to (3) are jointly significant when estimated in a fully-interacted model. Significance levels: *** p<0.01, ** p<0.05, * p<0.10

Appendix B Table 2b: Incentivized agreeableness regressions

	Treatment				p-value of (1)=(2)=(3)=0 (5)
	Extroversion (1)	Introversion (2)	No Priming (3)	Audit (4)	
Non-incentivized agreeableness	0.567*** (0.078)	0.714*** (0.112)	0.484*** (0.161)	0.762*** (0.081)	0.000
IQ	-0.036 (0.067)	0.069 (0.070)	0.127* (0.072)	0.034 (0.052)	0.232
Risk aversion	0.109 (0.077)	0.004 (0.063)	0.176** (0.070)	-0.053 (0.059)	0.040
Locus of control	0.033 (0.074)	0.110 (0.071)	0.101 (0.064)	0.079 (0.056)	0.168
Acceptability of lies for gain	0.017 (0.071)	0.108 (0.085)	0.053 (0.092)	0.091 (0.082)	0.577
Self-deception	-0.016 (0.074)	-0.071 (0.100)	-0.023 (0.100)	0.017 (0.075)	0.896
Impression management	-0.063 (0.077)	-0.187** (0.093)	0.010 (0.095)	-0.056 (0.072)	0.192
Self-monitoring	-0.051 (0.080)	0.093 (0.110)	-0.062 (0.100)	-0.100 (0.079)	0.680
Machiavellianism	0.029 (0.095)	0.092 (0.101)	0.081 (0.097)	0.014 (0.065)	0.656
Optimism	0.106 (0.077)	-0.086 (0.070)	0.003 (0.074)	-0.019 (0.064)	0.338
R ²	0.326	0.375	0.294	0.531	

Notes: Columns (1) to (4) report coefficient estimates from separate OLS regressions for each treatment. The dependent variable in each regression is the incentivized agreeableness score from the Evaluation session standardized using the mean and standard deviation of the incentivized scores from subjects in the same treatment. The non-incentivized agreeableness score used in the controls is standardized using the mean and standard deviation for the non-incentivized scores from subjects in the same treatment. The remaining controls—all of which were measured in the Baseline session except for locus of control—are standardized using the full sample of all subjects in all treatments. The standard errors reported in parentheses are scaled by the square of the observation’s variance estimate (i.e., the diagonal element of the projection matrix) as suggested by Davidson and MacKinnon [1993] and implemented in Stata using the “vce(hc3)” option. Column (5) reports the p-value for a heteroscedasticity-robust Wald test for the hypotheses that the coefficients in Columns (1) to (3) are jointly significant when estimated in a fully-interacted model. Significance levels: *** p<0.01, ** p<0.05, * p<0.10

Appendix B Table 2c: Incentivized openness regressions

	Treatment				p-value of (1)=(2)=(3)=0 (5)
	Extroversion	Introversion	No Priming	Audit	
	(1)	(2)	(3)	(4)	(5)
Non-incentivized openness	0.427*** (0.105)	0.602*** (0.080)	0.575*** (0.100)	0.768*** (0.061)	0.000
IQ	0.163** (0.071)	-0.090 (0.066)	0.041 (0.069)	0.011 (0.065)	0.060
Risk aversion	0.124 (0.077)	0.029 (0.079)	0.026 (0.062)	0.061 (0.051)	0.404
Locus of control	-0.063 (0.073)	0.163** (0.077)	0.044 (0.073)	-0.022 (0.049)	0.134
Acceptability of lies for gain	-0.011 (0.094)	0.059 (0.076)	-0.032 (0.079)	0.044 (0.056)	0.854
Self-deception	-0.048 (0.094)	0.028 (0.089)	-0.065 (0.092)	0.067 (0.067)	0.838
Impression management	-0.024 (0.096)	-0.013 (0.085)	-0.004 (0.088)	0.005 (0.058)	0.994
Self-monitoring	0.120 (0.077)	-0.017 (0.100)	0.048 (0.094)	0.039 (0.078)	0.437
Machiavellianism	-0.034 (0.101)	0.069 (0.106)	0.015 (0.090)	-0.004 (0.084)	0.905
Optimism	0.015 (0.080)	-0.128 (0.081)	-0.046 (0.082)	-0.059 (0.075)	0.416
R ²	0.233	0.442	0.334	0.617	

Notes: Columns (1) to (4) report coefficient estimates from separate OLS regressions for each treatment. The dependent variable in each regression is the incentivized openness score from the Evaluation session standardized using the mean and standard deviation of the incentivized scores from subjects in the same treatment. The non-incentivized openness score used in the controls is standardized using the mean and standard deviation for the non-incentivized scores from subjects in the same treatment. The remaining controls—all of which were measured in the Baseline session except for locus of control—are standardized using the full sample of all subjects in all treatments. The standard errors reported in parentheses are scaled by the square of the observation's variance estimate (i.e., the diagonal element of the projection matrix) as suggested by Davidson and MacKinnon [1993] and implemented in Stata using the “vce(hc3)” option. Column (5) reports the p-value for a heteroscedasticity-robust Wald test for the hypotheses that the coefficients in Columns (1) to (3) are jointly significant when estimated in a fully-interacted model. Significance levels: *** p<0.01, ** p<0.05, * p<0.10

Appendix B Table 2d: Incentivized neuroticism regressions

	Treatment				p-value of (1)=(2)=(3)=0 (5)
	Extroversion (1)	Introversion (2)	No Priming (3)	Audit (4)	
Non-incentivized neuroticism	0.243** (0.107)	0.426*** (0.084)	0.432*** (0.111)	0.669*** (0.084)	0.000
IQ	-0.143** (0.072)	0.040 (0.070)	-0.047 (0.076)	-0.060 (0.062)	0.201
Risk aversion	-0.121 (0.086)	-0.070 (0.086)	-0.047 (0.073)	0.012 (0.058)	0.367
Locus of control	0.018 (0.081)	-0.231*** (0.072)	-0.131 (0.083)	0.015 (0.068)	0.006
Acceptability of lies for gain	-0.086 (0.094)	-0.117 (0.081)	0.035 (0.103)	-0.116 (0.089)	0.381
Self-deception	-0.025 (0.086)	-0.009 (0.096)	0.012 (0.098)	-0.088 (0.075)	0.991
Impression management	-0.062 (0.095)	0.122 (0.097)	0.012 (0.099)	0.064 (0.079)	0.569
Self-monitoring	-0.030 (0.093)	0.113 (0.104)	0.013 (0.106)	0.004 (0.105)	0.726
Machiavellianism	0.015 (0.097)	-0.197* (0.113)	-0.085 (0.105)	-0.004 (0.108)	0.292
Optimism	-0.084 (0.093)	0.073 (0.085)	0.049 (0.104)	0.097 (0.095)	0.623
R ²	0.137	0.302	0.220	0.444	

Notes: Columns (1) to (4) report coefficient estimates from separate OLS regressions for each treatment. The dependent variable in each regression is the incentivized neuroticism score from the Evaluation session standardized using the mean and standard deviation of the incentivized scores from subjects in the same treatment. The non-incentivized neuroticism score used in the controls is standardized using the mean and standard deviation for the non-incentivized scores from subjects in the same treatment. The remaining controls—all of which were measured in the Baseline session except for locus of control—are standardized using the full sample of all subjects in all treatments. The standard errors reported in parentheses are scaled by the square of the observation’s variance estimate (i.e., the diagonal element of the projection matrix) as suggested by Davidson and MacKinnon [1993] and implemented in Stata using the “vce(hc3)” option. Column (5) reports the p-value for a heteroscedasticity-robust Wald test for the hypotheses that the coefficients in Columns (1) to (3) are jointly significant when estimated in a fully-interacted model. Significance levels: *** p<0.01, ** p<0.05, * p<0.10

Appendix B Table 3: Robustness checks for extroversion regressions

	Treatment			
	<i>Extroversion</i>	<i>Introversion</i>	<i>No Priming</i>	<i>Audit</i>
Controls	(1)	(2)	(3)	(4)
Non-incentivized extroversion	0.501*** (0.144)	0.511*** (0.117)	0.376*** (0.121)	0.644*** (0.086)
IQ	0.146** (0.072)	-0.147** (0.071)	0.089 (0.083)	-0.001 (0.066)
Risk aversion	0.161** (0.070)	0.067 (0.092)	0.009 (0.077)	-0.033 (0.059)
Locus of control	-0.076 (0.089)	0.043 (0.077)	0.115 (0.084)	0.026 (0.067)
Acceptability of lies for gain	-0.008 (0.110)	0.082 (0.087)	-0.018 (0.114)	0.050 (0.080)
Self-deception	0.080 (0.088)	-0.005 (0.096)	-0.033 (0.109)	0.108 (0.076)
Impression management	-0.201* (0.110)	0.126 (0.107)	0.016 (0.112)	-0.101 (0.079)
Self-monitoring	-0.159 (0.097)	-0.080 (0.128)	0.167* (0.095)	0.051 (0.091)
Machiavellianism	-0.021 (0.103)	0.013 (0.127)	0.020 (0.097)	0.013 (0.098)
Optimism	0.087 (0.096)	-0.040 (0.100)	0.010 (0.096)	-0.099 (0.082)
Constant	-0.005 (0.078)	0.007 (0.075)	-0.013 (0.079)	0.029 (0.062)
R ²	0.253	0.270	0.297	0.508
# of subjects	153	152	137	158

Notes: Columns (1) to (4) report coefficient estimates from specifications identical to those in Table 4 using incentivized extroversion as the dependent variable when the sample is restricted to subjects who provided that same answer to all five Big 5 items on the same screen zero or one time out of 20 screens. Columns (5) to (8) report the same estimates restricting the sample to subjects who give no contradictory responses in pairs of four closely related items. Column (9) reports estimates for the *Extroversion* treatment using the censored least absolute deviations estimator to account for the right-censoring of six observations using the “clad” command in Stata.

Appendix B Table 3 continued: Robustness checks for extroversion regressions

	Treatment				
	<i>Extroversion</i>	<i>Introversion</i>	<i>No Priming</i>	<i>Audit</i>	<i>Extroversion</i>
Controls	(5)	(6)	(7)	(8)	(9)
Non-incentivized extroversion	0.423*** (0.143)	0.456*** (0.116)	0.387*** (0.117)	0.662*** (0.095)	0.291 (0.175)
IQ	0.191** (0.078)	-0.182*** (0.069)	0.055 (0.084)	0.006 (0.069)	0.260 (0.098)
Risk aversion	0.132* (0.073)	0.089 (0.089)	0.009 (0.075)	-0.034 (0.059)	0.161 (0.106)
Locus of control	-0.028 (0.085)	0.060 (0.082)	0.144* (0.083)	0.017 (0.067)	-0.001 (0.119)
Acceptability of lies for gain	-0.010 (0.108)	0.041 (0.093)	-0.017 (0.123)	0.039 (0.085)	0.082 (0.138)
Self-deception	0.052 (0.083)	0.029 (0.098)	-0.051 (0.104)	0.105 (0.078)	0.014 (0.102)
Impression management	-0.160 (0.111)	0.112 (0.105)	-0.010 (0.111)	-0.101 (0.082)	-0.155 (0.113)
Self-monitoring	-0.087 (0.090)	-0.025 (0.124)	0.143 (0.097)	0.057 (0.106)	-0.090 (0.119)
Machiavellianism	-0.005 (0.101)	-0.036 (0.133)	0.003 (0.097)	0.018 (0.112)	0.015 (0.135)
Optimism	0.125 (0.092)	-0.008 (0.092)	0.051 (0.095)	-0.141* (0.083)	0.126 (0.123)
Constant	-0.015 (0.075)	0.000 (0.075)	0.020 (0.078)	0.015 (0.063)	0.142 (0.111)
R ²	0.238	0.263	0.296	0.503	
# of subjects	157	155	135	153	167

Notes: Columns (1) to (4) report coefficient estimates from specifications identical to those in Table 4 using incentivized extroversion as the dependent variable when the sample is restricted to subjects who provided that same answer to all five Big 5 items on the same screen zero or one time out of 20 screens. Columns (5) to (8) report the same estimates restricting the sample to subjects who give no contradictory responses in pairs of four closely related items. Column (9) reports estimates for the *Extroversion* treatment using the censored least absolute deviations estimator to account for the right-censoring of six observations using the “clad” command in Stata with bootstrapped standard errors in parentheses.

Appendix B Table 4: Personality trait extreme adjective lists

Extroversion:
active ad-lib adventuresome adventurous affable aggressive alert alive anxious approachable ardent articulate assertive astir at-work audacious avid bold bustling busy casual chatty close clubby concerned confident convivial cordial courageous daring decisive determined diligent dynamic eager earnest effective efficacious effusive emphatic energetic engaged enterprising enthusiastic excessively communicative exertive extraverted exuberant familiar fanatical fearless fervent flowing forceful friendly full-of-life functioning garrulous genial glib going good-natured gregarious hasty heroic impelling impromptu in-force in-play in-process industrious insistent instinctive intense interested intimate intrepid keen lively loquacious mobile movable moving neighborly offhand operating operative outgoing passionate pleased powerful progressive pushing pushy quick rabid rapid ready regular resolute restless rhapsodic risk-taking risky rolling running rushing rustling self-assured self-confident shifting simmering simple sociable social speeding spirited spontaneous sprightly spry streaming strong talkative tireless traveling turning unplanned vigorous voluble voluntary walking warm willing working zealous
Introversion:
aloof ambivalent antagonistic apathetic apathetic bashful blasé brooding buttoned-up careless cautious cold collected confused conservative constrained cool coy demure dependable diffident distant dormant dull easy-going embarrassed emotionless fearful feeble frightened frustrated gentle guarded half-hearted hostile humble hushed idle immobile inaccessible inactive inattentive inert inhibited inhospitable inoperative introspective introverted jobless lackadaisical laid-back lazy lethargic modest mum mute nervous non-gregarious passionless passive prudent quiet realistic recessive reclusive reliable repressed restrained reticent retiring secretive sedentary self-conscious sensitive sheepish shy silent sleepy sluggish soft-spoken standoffish static stuck-up subdued sure tepid timid tired unadventurous unapproachable unassertive unbending uncommunicative uncongenial unemployed unenergetic unenthusiastic unexcited unforthcoming unfriendly uninterested unneighborly unsociable unsocial uppity weak weary withdrawn
Conscientiousness:
accountable accurate all-encompassing answerable arranged assiduous businesslike careful cautious circumspect complete comprehensive conscientious constructive coordinated culpable decent decisive dedicated deliberate dependable detailed diligent discreet down-to-earth economical efficient exhaustive factual fastidious feasible formed formulated frugal full functional fussy genuine good guarded guilty hardworking honest humorless important in-charge intensive judicious leery liable meticulous organized painstaking persevering pledged positive possible practicable practical pragmatic predictable profound prudent rational realistic reasonable reliable respectable responsible safe sane sensible serious severe sincere sober solid somber stable standardized steady stingy strong subject systematized tentative thorough thoughtful thrifty trustworthy useful vigilant wary watchful workable
Non-conscientiousness:
absurd apathetic audacious bizarre bold capricious careless casual chaotic confused costly daring disorganized dull exaggerated excessive exorbitant expensive extravagant fanciful fancy fantastic flamboyant flashy foolhardy foolish forgetful frivolous garish gaudy grandiose haphazard hasty idiotic ill-advised ill-considered illogical immature immoral impetuous implausible impossible impractical improbable imprudent impulsive inaccurate inadvertent inattentive indifferent indiscreet indulgent irrational irresponsible lackadaisical lavish lax lazy lethargic ludicrous messed-up muddled negligent nonchalant ornate ostentatious outrageous overpriced passive petty pointless profligate quixotic rash reckless ridiculous senseless showy silly sleepy sloppy sluggish speculative thoughtless tired trivial unattainable undependable unmethodical unreal unrealistic unreliable unusable unwise unworkable useless wasteful weary
Agreeableness:
acceptable acquiescent adaptable affable affectionate agreeable altruistic amenable amiable attentive benevolent biddable big big-hearted buoyant candid charitable cheerful civil civilized clean comfortable compassionate complaisant compliant conciliatory confident congenial consenting considerate cooperative cordial courteous courtly decent deferential diplomatic discretionary docile eager enthusiastic equal equitable fair favorable flexible friendly generous genteel gentle giving good good-natured gracious happy heartfelt hearty helpful honest honorable hospitable humane humanitarian impartial kind kindhearted kindly lavish lawful legitimate loving mannerly neighborly noble objective optimistic pleasant polite principled proper reasonable respectful responsive rosy sanguine self-effacing selfless sincere sociable straightforward supportive sympathetic tender thoughtful tolerant trustful trusting trustworthy unbiased understanding unprejudiced unselfish upbeat useful warm warmhearted well-behaved well-mannered willing
Disagreeableness:
abrupt abusive adamant aloof apathetic arbitrary bad-mannered bad-tempered bellicose biased blunt boorish brusque cantankerous cautious churlish coarse cold cold-blooded contrary cool crude cruel dead disagreeable disbelieving discriminatory dishonest disrespectful distant distrustful doubtful dubious egocentric egoistic egotistical egomaniacal egotistic egotistical emotionless fearful frigid frosty frugal greedy harsh hateful icy ignorant illegal immoral immutable impolite improper inequitable inexcusable inflexible inhospitable insensitive insulting intractable intrusive irritable jealous leery lukewarm malicious miserly mistrustful narcissistic nasty obnoxious obscene obstinate one-sided partisan penny-pinching perverse prejudiced resolute rigid rigorous rough rude self-centered selfish shameful skeptical spiteful steadfast stingy strict stringent stubborn surly suspicious thinks-only-of-oneseif thrifty uncaring uncivil uncompromising uncooperative uneasy unethical unfair unfriendly ungenerous unjust unjustifiable unkind unlawful unpleasant unreasonable unresponsive unwarranted unyielding vulgar wary wrong wrongful
Openness:
accomplished advanced alert analytic analytical artistic astute brainy bright brilliant civilized classy cogent contemplative cosmopolitan creative cultivated cultured curious deliberate delicate desiring-knowledge detailed discerning discerning discreet discriminating educated elegant enlightened experienced extricated extraneous fanciful fantastic genteel gifted imaginative incisive informed ingenious innovative inquisitive insightful intelligent interested interpretive intuitive inventive investigative knowing knowledgeable lettered literate mature meditative observant offbeat open original originitive penetrating perceptive polished polite practical precise productive prolific rational refined reflective resourceful romantic scientific sensitive smart sophisticated studious subtle systematic tasteful thorough thoughtful tolerant traveled unexcited urbane versed visionary vivid well-bred well-informed whimsical wise worldly
Non-openness:
absent-minded antiseptic arid artless aseptic banal barbaric bare barren bleak boorish brainless brutish careless casual coarse crass crude cursory dead decontaminated deficient dense desert desolate disinfected doltish dry dull dumb earthy easily-pleased effete empty empty-headed fallow foolish forgetful fruitless futile gaunt germ-free gross guileless hackneyed half-witted hasty hygienic idiotic ignorant imbecilic imperceptive impotent imprecise inaccurate indifferent indiscreet inelegant inexperienced infecund infertile irresponsible lackadaisical lax naive natural negligent nonchalant pasteurized perfunctory philistine predictable primitive reckless rude sanitary septic simple sloppy sterile sterilized straightforward stupid tedious thoughtless trite unanalytical uncivilized uncouth uncreative uncultured unfruitful unfussy unimaginative uninfected uninquisitive uninspired unintelligent uninventive uninvolved unlettered unoriginal unpolished unrefined unreflective unsophisticated wasteful
Neuroticism:
affecting afraid agitated ambiguous angry annoyed anxious apologetic apprehensive ashamed bitching bitter blue cantankerous changeable close complaining concerned contrite crabby cross demonstrative-about-feelings disaffected discontented disgruntled displeased disquieted dissatisfied distrustful disturbed doubtful downcast edgy emotional enraged envious erratic exasperated exciting fearful fickle firm furious fussy greedy griping guilt-ridden guilty heated hesitant hysterical impassioned indignant insecure irate irrational irritable irritated jealous jittery jumpy kvetching mad melancholy moody moving nervous neurotic offended outraged passionate perturbed poignant precarious regretful repentant resentful risky rocky sad sensitive sentimental shaky shy skittish slippery sorrowful sorry spontaneous spooked stiff strained stretched sulky sullen suspicious temperamental tense ticklish tight timid touching touchy tricky troubled uncertain uneasy unhappy unpredictable unsettled unstable unsteady unsure upset uptight vexed volatile watchful weak wobbly worried
Emotional stability:
abiding able above-suspicion assured at-ease balanced breezy calm carefree casual certain cheerful cheery clean clean-handed clear collected composed confident constant content contented cool crimeless deadpan defended definite determined dispassionate durable easy easygoing emotionless enduring established even exemplary fast faultless firm fixed flexible glad good gratified guarded guilt-free happy happy-go-lucky harmonious impassive informal jaunty jovial laid-back lasting lighthearted low-key mild moderate nonchalant non-envious non-responsible not responsive peaceful permanent placid pleased poised protected quiet reassured regular relaxed reliable reticent safe satisfied secure serene settled sheltered shielded slow smooth solid soothing sound spontaneous stable steady strong substantial sure thankful tolerant tranquil unemotional untroubled