Intentional Response Distortion on Personality Tests: Using Eye-Tracking to Understand Response Processes When Faking

Edwin A. J. van Hooft
University of Amsterdam

Marise Ph. Born
Erasmus University Rotterdam

Intentional response distortion or faking among job applicants completing measures such as personality and integrity tests continues to be a concern in personnel selection. The present study aimed to investigate whether eye-tracking technology can improve our understanding of the response process when faking. In an experimental within-participants design, a Big Five personality test and an integrity measure were administered to 129 university students in 2 conditions: a respond honestly and a faking good instruction. Item responses, response latencies, and eye movements were measured. Results demonstrated that all personality dimensions were fakeable. In support of the theoretical position that faking involves a less cognitively demanding process than responding honestly, we found that response times were on average 0.25 s slower and participants had less eye fixations in the fake good condition. However, in the fake good condition, participants had more fixations on the 2 extreme response options of the 5-point answering scale, and they fixated on these more directly after having read the question. These findings support the idea that faking leads to semantic rather than self-referenced item interpretations. Eye-tracking was demonstrated to be potentially useful in detecting faking behavior, improving detecting rates over and beyond response extremity and latency metrics.

Keywords: faking, eye-tracking, response latency, personality test, integrity

This article was published Online First October 3, 2011.
Edwin A. J. van Hooft, Work and Organizational Psychology, University of Amsterdam, Amsterdam, the Netherlands; Marise Ph. Born, Institute of Psychology, Erasmus University Rotterdam, Rotterdam, the Netherlands.

A previous version of this article was presented at the 24th Annual Conference of the Society for Industrial and Organizational Psychology, April 2009, New Orleans, Louisiana. This research was supported in part by an innovation grant from SenterNovem (Nr. INV050701). We thank Marcel L’Herminez (i-Test and Selection Center) for his help with planning and organizing the study, Annemarie Renger (i-Test and Selection Center) and the employees of the Rotterdam office of Verify Netherlands for their help with the data collection, Erwin Maas (Verify Netherlands) for preparing the data files, and Conor Dolan for his help with data analysis.

Correspondence concerning this article should be addressed to Edwin A. J. van Hooft, University of Amsterdam, Work and Organizational Psychology, Weesperplein 4, 1018 XA Amsterdam, the Netherlands. E-mail: e.a.j.vanhoof@uva.nl

Meta-analytic work demonstrated that individuals are able to fake on self-report non-cognitive measures when instructed to do so (Alliger & Dwight, 2000; Viswesvaran & Ones, 1999). Furthermore, although there has been substantial debate in the literature, meta-analytic work on differences between job applicants’ and job incumbents’ scores demonstrates that it is likely that faking does occur among actual job applicants (Birkeland et al., 2006). The proportion of fakers in applicant samples varies between approximately 20% and 50%. For example, in a study using randomized response techniques, more than 30% of recent applicants admitted that they had engaged in faking behavior (Donovan, Dwight, & Hurtz, 2003). Griffith, Chmielowski, and Yoshita (2007) had recent job applicants complete the same personality test again under an honest instruction, and they found that (depending on the estimation method used) between 22% and 49% had elevated their scores during the job application.

As individuals can and do fake on self-report personality and integrity tests, an important issue then is whether faking poses a threat to the validity of such tests. Several studies indicated that faking does not impact the validity (e.g., Barrick & Mount, 1996; Ellingson, Smith, & Sackett, 2001; Hough, 1998; Hough, Eaton, Dunnette, Kamp, & McCly, 1990; Li & Bagger, 2006; Ones & Viswesvaran, 1998; Ones, Viswesvaran, & Reiss, 1996; Smith & Ellingson, 2002). These studies, however, almost exclusively operationalized faking using social desirability, impression manage-
ment, or lie scale scores. Because such self-report scales to assess faking themselves have been found to be sensitive to faking (e.g., Viswesvaran & Ones, 1999; Zickar & Robie, 1999) and because their validity has been questioned (Griffith et al., 2006; Griffith & Peterson, 2008; Stark et al., 2001; see also McFarland & Ryan, 2000), it may be problematic to draw firm conclusions about the effects of faking from studies using such scales (Burns & Christiansen, 2006). Indeed, studies using other paradigms such as comparing applicants with non-applicants (e.g., Griffin, Hesketh, & Grayson, 2004; Rosse, Stecher, Miller, & Levin, 1998; Schmit & Ryan, 1993; Stark et al., 2001) or comparing scores in a faking condition with scores in a responding honestly condition (e.g., Douglas, McDaniel, & Snell, 1996; Holden, Wood, & Tomashewski, 2001; McFarland & Ryan, 2000; Zickar & Robie, 1999) typically conclude that faking does impact both the construct and criterion-related validity. Regarding the construct validity, for example, faking has been found to affect the factor structure of personality measures (Schmit & Ryan, 1993; Zickar & Robie, 1999) and to lead to differential item functioning (Griffin et al., 2004; Stark et al., 2001). Regarding criterion-related validity, studies comparing job applicants with job incumbents (Rosse et al., 1998), comparing a faking and a responding honestly condition (Douglas et al., 1996; Griffith et al., 2007; Holden et al., 2001), and Monte Carlo simulation studies (Komar et al., 2008) have demonstrated that faking negatively impacts the criterion-related validity and affects hiring decisions based on personality tests.

If faking on personality and integrity tests indeed impacts the construct validity, criterion-related validity, and hiring decisions, this raises the issue of how to detect faking in a selection setting. In the selection practice, testing companies usually include social desirability, impression management, or lie scales to identify fakers. Given the finding that such measures may have low validity, techniques that do not rely on self-report measures may be more useful for detecting faking behavior. Based on this reasoning, the purpose of the present study was twofold. First, we aimed to investigate whether eye-tracking technology can increase our understanding of the response processes when faking compared to answering honestly. When studying the effects of faking on score elevations, Viswesvaran and Ones (1999) recommended using a within-participant design because such a design removes possible effects of individual differences in faking propensity, resulting in more accurate estimates of score elevations when faking. Similarly, substantial individual differences have been reported in response latencies (Holden, Fekken, & Cotton, 1991; Holden, Kroner, Fekken, & Popham, 1992) and eye behavior (Rayner, 1998). By comparing response latencies and eye behavior between faking and honest responding within participants, the effects of such between-individual differences are removed, providing more fine-grained information about differences between the two conditions in underlying response processes. Regarding our second purpose (i.e., exploring whether eye-tracking technology may yield information that can be used for identifying faking behavior), a between-participants design was deemed preferable, because such a design more accurately represents actual test-taking settings (e.g., applicants taking a personality test during a selection procedure).

**Faking and Scores on Personality Factors and Integrity**

Meta-analyses have reported evidence for faking on personality and integrity measures across different study designs. For example, in a meta-analysis of experimental studies, Viswesvaran and Ones (1999) reported medium to large effect sizes for mean differences on the Big Five factors between faking and honest conditions, with $d$s ranging from 0.48 to 0.65 in between-participants designs, and from 0.47 to 0.93 in within-participants designs. Alliger and Dwight (2000) reported effect sizes between 0.59 and 1.02 in a meta-analysis of studies on integrity tests using between-participants designs. In a meta-analysis of studies comparing job applicants’ and non-applicants’ personality scores, Birkland et al. (2006) found small to medium mean differences ($d$s between 0.11 and 0.45). Although response distortion occurs on all Big Five factors and integrity, these meta-analyses reported the largest effect sizes for Neuroticism and Conscientiousness. The finding that people inflate their scores the most on Neuroticism...
Faking and Response Latencies

Several studies have examined whether faking is associated with response latencies when answering personality items. An item’s response latency reflects the time that elapses between the presentation of the item and the occurrence of a response to the item (cf. Hsu, Santelli, & Hsu, 1989). Theory on the effects of faking on response latencies is still unclear (e.g., Fluckinger, McDaniel, & Whetzel, 2008), and several contrasting theoretical perspectives have been proposed (Holden et al., 2001; Holtgraves, 2004; Vasilopoulos, Reilly, & Leaman, 2000). These theoretical perspectives build upon stage models of the response process. That is, when answering to self-report items, theory suggests that respondents move through the following stages (Tourangeau & Rasinski, 1988; see also Holtgraves, 2004; McDaniel & Timm, 1990): (a) interpretation of the item, (b) retrieval of relevant information, (c) rendering a judgment based on the retrieved information, and (d) mapping the judgment onto the format of the answer scale and executing the response.

A first theoretical perspective suggests that faking takes time, leading to longer response latencies. Previous research on lying and deceiving, for example, has conceptualized lying as a cognitively more complex task than telling the truth (e.g., Zuckerman, DePaulo, & Rosenthal, 1981), resulting in a higher cognitive load (Vrij, Edward, & Bull, 2001). In their comprehensive meta-analysis, DePaulo et al. (2003) reported that lying is generally associated with increased response latencies when people cannot prepare their answers. Tourangeau and Rasinski (1988) offered a rationale for longer response times when faking by theorizing that answers may undergo an editing process in which the answer is checked for social desirability. Similarly, McDaniel and Timm (1990) suggested that compared to telling the truth, lying on an item or faking should take time because it involves an additional stage in the response process, that is, formulating the decision to lie. Consistent with these ideas, McDaniel and Timm found that dishonest responses on a biodata instrument took on average 0.6 s longer than honest responses. Similarly, in three experimental studies, Holtgraves (2004) found that heightened concerns with social desirability resulted in longer response times.

A second, opposing, theoretical perspective is that faking causes shorter response latencies. Hsu et al. (1989) and Holtgraves (2004) noted that the process of faking may involve more primitive cognitive processing than honest responding, suggesting that respondents do not move through all stages of the response process. More specifically, Hsu et al. stated that whereas honest responding leads to a self-referenced interpretation of the item content, faking leads to a purely semantic item interpretation, which takes less processing time. Holtgraves noted that social desirability responding may be characterized by the less complex response process of direct retrieval, meaning that when faking, respondents do not try to retrieve accurate information but produce a response based solely on the fake instruction and the social desirability of the item. Consistent with this perspective, Hsu et al. reported faster response times when faking than when responding honestly. Furthermore, Holden, Fekken, and Jackson (1985) found a negative correlation between response latency and social desirability of personality items, indicating that socially desirable items take less time to respond to. Also in support of the notion that faking is faster, Holden et al. (2001) showed that limiting respondents’ answering time did not prevent people from faking.

A third perspective, described by Holden et al. (1992), states that the effects of faking on response latencies depend on the faking schema and the social (un)desirability of the test item. Specifically, when respondents adopt a faking good strategy, socially desirable items align with the adopted schema (i.e., creating a favorable impression) and are therefore easy to respond to. Socially undesirable items, however, are inconsistent with a faking good schema, as they are indicative of creating a non-favorable impression, and therefore more difficult to respond to. Thus, compared to honest responding, faking good leads to faster responses on socially desirable items and slower responses on socially undesirable items (Holden et al., 2001). Several studies have found support for this interactive model of faking (e.g., Brunetti, Schloßmann, Scott, & Hollrah, 1998; Holden & Kroner, 1992; Holden et al., 1992).

In the present study, we examined which theoretical perspective is supported by our data. If faking is more cognitive complex and adds an extra stage to the response process (e.g., deciding to fake, response editing), response latencies should be larger in the fake good than in the honest condition. If faking involves a more primitive response process (i.e., semantic rather than self-referenced item interpretation, direct retrieval), response latencies should be larger in the honest than in the fake good condition. If Holden et al.’s (1992) interactive model is valid, participants’ response latencies should be lower in the fake good than in the honest condition for positively keyed items, and higher in the fake good than in the honest condition for negatively keyed items.

Faking and Eye Behavior

An important contribution of the present study is the investigation of the effects of faking on eye behavior when responding to test items. When reading or looking at a picture, our eyes alternately make rapid movements, called saccades, or remain relatively still during fixations on specific regions of interest (Rayner, 1998). Saccades last about 15–40 ms and mainly serve to move the eye from one fixation point to the next (Reichle, Polatsek, Fisher, & Rayner, 1998). Information is largely obtained during fixations, because the eyes are moving too quickly during saccades (Rayner, 1998). Therefore, analyses of eye behavior are usually based on eye fixations (Reichle et al., 1998). In the present study, we focus on the number and the location of the eye fixations, proposing that such data may provide valuable information about the response process when faking, beyond response latency data. That is, fixation data allow for examining what areas of the test item (i.e., the question, neutral or extreme response options) draw the attention
of the test-taker, providing information about the underlying cognitive processes.

Because, to our knowledge, no research or theory has addressed eye behavior in the context of faking on personality tests, we resorted to general research on lying and deceiving (see DePaulo et al., 2003), work on eye behavior in reading and information processing (see Rayner, 1998), and theory and research on faking and response latencies to formulate tentative hypotheses regarding the effects of faking on eye behavior. First, based on the same rationales as for response latencies, if lying or faking is cognitively more complex than answering honestly (i.e., containing an extra stage in the response process), the number of eye fixations is expected to be larger in the fake good than in the honest condition. That is, previous research on eye behavior when reading demonstrated that cognitive load relates to an increase in eye fixations (Rayner, 1998). Second, if lying or faking is associated with a more primitive response process than telling the truth (i.e., direct retrieval, semantic item interpretations), cognitive load is lower, which should result in less fixations in the faking good than in the honest condition. Third, extending Holden et al.'s (1992) interactive model to eye fixations, the number of fixations should be lower in the fake good than in the honest condition for positively keyed items, and higher in the fake good than in the honest condition for negatively keyed items.

In addition to the number of eye fixations during item responding, we examined (a) the location of the fixations and (b) the order of fixation locations (i.e., eye paths). When faking good, participants aim at getting higher total scores on the personality factors and integrity (see Alliger & Dwight, 2000; Birkeland et al., 2006; Viswesvaran & Ones, 1999). Because higher scores result from giving more extreme responses, in the faking good condition, supposedly more attention is paid to the extreme response categories than to the middle response categories. Consequently, in the faking good condition, more fixations are expected on the extreme response categories that reflect a high score (which differs for positively and negatively keyed items). In the honest condition, in contrast, more eye fixations are expected on the middle response categories. We also explored the eye paths when reading and responding to test items. As mentioned, when reading, the eyes alternate focus on points of interest during fixations and move between these points during saccades. Eye paths can be construed based on the sequence of fixation locations. Such eye paths may be reflective of the underlying response process when faking or answering honestly. For example, a purely semantic item interpretation and corresponding socially desirable response may result in a more direct eye path such that after fixating on the question, the eyes directly move to the socially desirable response option. Based on this reasoning, we explored whether the eye paths differed between the faking good and honest condition. If faking good leads to a more primitive response process indicated by a semantic rather than self-referenced item interpretation, more eye paths in the form of question → extreme response option are to be expected.

**Differentiating Between Honest Responding and Faking**

If faking leads to different eye behavior compared to responding honestly, then an interesting question is whether such differences may yield information that is useful to detect faking behavior. To explore this issue, we construed metrics that differentiated between faking and responding honestly, and we used these to examine how well faking can be identified. Because previous research on mean differences (i.e., Alliger & Dwight, 2000; Birkeland et al., 2006; Viswesvaran & Ones, 1999) suggests that when faking good participants give more extreme responses, a first metric reflected the proportion of extreme responses. Further, previous research has demonstrated that response latencies can be used to differentiate between honest and faking responses (e.g., Holden, 1998; Holden & Hibbs, 1995; Holden & Kroner, 1992). For example, in his experimental study, Holden (1998) found that response latency measures were useful to distinguish fakers from non-fakers with a hit rate of 64.5%. Other studies using response latency measures to differentiate between a fake and honest condition found classification hit rates varying between 62% and 82% (see Holden & Hibbs, 1995). Therefore, a second cluster of metrics was based on response latencies. In the present study, we sought to extend this research by exploring whether a third cluster of metrics based on eye-tracking data (i.e., number, location, and order of eye fixations) can improve the identification of faking beyond the use of extreme response and latency data.

**Method**

**Participants and Design**

Participants were 129 university students (81 women and 48 men; mean age = 21.91 years, SD = 3.52). Students participated for either course credit or €15. A 2 × 2 mixed design was used with response instruction (i.e., answer honestly vs. fake good) as the within-participants factor and order of instruction as the between-participants factor. For the within-participants factor, respondents were asked to respond to 105 items on personality and integrity, once honestly and once faking good. Order of instruction was randomized.

**Procedure, Materials, and Manipulations**

Data were collected in the eye-tracker laboratory of a market research company. This company specialized in eye-tracking research for commercial and scientific purposes (e.g., Pieters & Wedel, 2004). Upon entering the laboratory, participants were registered and given an identification card with a unique participant number. Participants were told that the total test session would take about 75 min and consisted of three parts: a first personality test, a cognitive ability test, and a second personality test. Participants were further explained that both personality tests would be administered at an eye-tracker. Although participants thus were aware that their eye movements were recorded, they were not aware of the purpose of the study and the eye-tracking. If applicable, participants were asked to remove eye mascara, hard contact lenses, and glasses to ensure the reliability of the eye-tracking. If needed, special lightweight replacement glasses were available.

After registration, the participants were seated behind one of the eye-trackers (see Figure 1). Instructions and test items were presented on a 21-in. (53.34-cm) liquid crystal display (LCD) touch screen monitor with a 1,280 × 1,024 pixel resolution, which was built into a table. Participants looked through a glass sheet to the
monitor. Participants were free to move their heads within a space of about 30 cm. Cameras (situated at the top of the eye-tracker) tracked the specific position at which the fovea was directed every 20 ms, using infrared corneal reflection (Duchovski, 2003). That is, the glass sheet between the participant and the LCD monitor reflected the infrared beam from the camera to the eye and back. Because infrared light is invisible to the eye, it does not distract participants. After calibration, this technology assesses the eye gaze position on the LCD screen with a measurement precision better than 0.5 degrees of visual angle (Pieters & Wedel, 2004).

For each participant, the session started with a calibration procedure to ensure measurement precision. Participants had to look at several dots spread on the screen, based on which the eye tracker was calibrated. After calibration, participants were presented with either the respond honestly or the fake good instruction. The instructions were adapted from McFarland and Ryan (2000). The respond honestly instruction read as follows:

In the next screens you will be presented with 105 questions with five response options. Please answer the questions as honestly as possible. Your answers will remain completely confidential and anonymous, and will be used for research purposes only. For this study we are interested in how you really are. Therefore it is very important that you answer the following questions as accurately and honestly as you can.

The fake good instruction read as follows:

Please imagine that you are graduated and are applying for a job. As part of the selection procedure you are presented with the following 105 questions with five response options. Please answer the questions such that you will come across as the ideal employee. For this study we are not interested in what your real answers for each question would be. Instead, for each question please select the answer that you feel will give you the best rank and make you look like the most suitable job applicant.

Following the instruction screen, each of the 105 items of the personality and integrity test was presented on a separate screen. The question was presented at the top, followed by the five response options in boxes. After having touched one of the boxes on the screen, the next item was presented. Answering all 105 items took about 10 min.

After having finished the first test on the eye-tracker, participants were brought to a different room and were seated behind a laptop computer to complete a cognitive ability test as a filler task. This test consisted of a short instruction, practice items, and four subtests, and it took about 45 min. Then participants were brought back to the eye-tracker room to complete the second administration of the 105 items with either the honest or the fake good instruction, depending on the instruction they had at the first administration. The second session was followed by a manipulation check item and two items on perceived task difficulty. Lastly, participants were debriefed, given their course credit or €15, and presented with a short report of their performance on the cognitive ability test.

Measures

Personality. The Five Factor Personality Inventory (FFPI; Hendriks, Hofstee, & De Raad, 1999)—consisting of 100 brief statements with a Likert response format ranging from 0 (much less [often] than others) to 4 (much more [often] than others)—was used to assess personality. Each of the five personality factors (i.e., Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Autonomy) were measured with 10 positively keyed items and 10 negatively keyed items. Items were recoded such that high scores reflect high levels on the factor. Positively and negatively keyed sample items are “I love to chat” and “I keep apart from others” for Extraversion, “I respect others’ feelings” and “I order people around” for Agreeableness, “I do things according to a plan” and “I make a mess of things” for Conscientiousness, “I readily overcome setbacks” and “I get overwhelmed by emotions” for Emotional Stability, and “I can easily link facts together” and “I follow the crowd” for Autonomy. The FFPI factors have been found to be reliable, stable, and of good construct validity, as demonstrated by high test–retest correlations (.74–.83) and self–other correlations (.54–.73; Hendriks et al., 1999). In further
support of the construct validity, De Fruyt, McCrae, Szirmai, and Nagy (2004) and Hendriks et al. (1999) found strong convergence with the NEO Personality Inventory—Revised (Costa & McCrae, 1992) across different countries for four of the FFPI factors (i.e., for Extraversion, rs ranged from .63 to .80; for Agreeableness, rs ranged from .53 to .81; for Conscientiousness, rs ranged from .49 to .81; and for Emotional Stability/Neuroticism, rs ranged from −.48 to −.83). Only for Autonomy the convergent correlations were somewhat lower (.20–.60). Table 1 displays the alpha coefficients.

**Integrity.** Integrity was measured with five items based on Wanek, Sackett, and Ones’s (2003) item-level review of integrity tests and Van Iddekinge, Taylor, and Eidson’s (2005) study on integrity facets. The items read as follows: “I am an honest person,” “I have occasionally stolen something small from another person” (reverse scored), “I have friends that are dishonest at times” (reverse scored), “Every now and then I have thought to take something that was not mine” (reverse scored), and “I am too honest to steal.” Response options were identical to those of the FFPI. Items were recoded such that high scores reflect high levels of integrity. See Table 1 for the alpha coefficients.

**Manipulation check and task difficulty items.** As a manipulation check, participants were asked the following question on a 5-point Likert scale: “You have answered to the same set of questions twice. To what extent did you give the same or different answers at both times?” The response options varied from 0 (completely the same) to 4 (completely different). Next, participants were given two items regarding the self-perceived task difficulty in the two conditions. Specifically, participants were asked the following: “How difficult or easy did you think it was to answer the questions as honestly as possible?” (cf. the honest condition) and “How difficult or easy did you think it was to answer the questions as coming across as the best job applicant?” (cf. the faking good condition). Response options varied from 0 (very difficult) to 4 (very easy) for both items.

**Response latencies.** For each participant, the response time in milliseconds was recorded for each test item in both conditions. The response time was defined as the time between the moment that the test item was presented on the screen and the moment that the participant touched the screen to answer that test item. Based on the item response times, the following average response latency scores were calculated. First, average response latencies were calculated for the Big Five factors and integrity per participant per condition by averaging the response latencies on the test items of each factor. Second, an average response latency total score for the honest condition and for the fake condition was calculated per participant by averaging all item response latencies per condition. Third, average response latency scores in the honest and fake condition were calculated for positively keyed and negatively keyed items separately.

**Eye-tracking measures.** An eye fixation is a period during which the eyes come to rest and pause at a region of interest (Reichle et al., 1998; Salvucci & Goldberg, 2000). In silent reading, fixation duration is 225 ms on average but may range from less than 100 ms to more than 500 ms depending on characteristics of the individual and the text (Rayner, 1998; Reichle et al., 1998). Several studies on reading have demonstrated that if readers are allowed 50–60 ms on each eye fixation, they read quite normally (e.g., Liversedge et al., 2004; Rayner, Inhoff, Morrison, Slowiczek, & Bertera, 1981). Based on this research, in the present study we included eye fixations (i.e., having the same gaze position at sequential measurement points) that lasted for at least three sequential measurement points of 20 ms. For each eye fixation, the exact pixel location on the LCD screen and the duration (in number of sequential measurement points of 20 ms) of the fixation was recorded. Furthermore, we defined several areas of interest on the screen for locating the fixation position. As shown in Figure 2, the question text (q), all five response options (a1–a5), and the total screen (d) were defined as boxes of interest. Based on the pixel location, it was recorded whether the fixation was within or outside each box of interest.

**Analyses and Results.**

The results on the manipulation check item indicated that 88.4% of the participants responded slightly to completely different in the two conditions, and 11.6% indicated that they had responded about the same. Because none of the participants indicated that they had responded completely the same in the two conditions, all participants were retained in subsequent analyses. Results on the task difficulty items showed that participants found faking more difficult (M = 1.91, SD = 1.09) than answering honestly (M = 2.37, SD = 0.82), t(128) = 4.14, p < .001.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Honest condition</th>
<th>Fake good condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>M</td>
</tr>
<tr>
<td>Extraversion</td>
<td>.93</td>
<td>2.50</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.81</td>
<td>2.60</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.90</td>
<td>2.29</td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>.91</td>
<td>2.67</td>
</tr>
<tr>
<td>Autonomy</td>
<td>.91</td>
<td>2.58</td>
</tr>
<tr>
<td>Integrity</td>
<td>.71</td>
<td>2.77</td>
</tr>
</tbody>
</table>

*Note. N = 129.
***p < .001.
Scale Scores on the Big Five Factors and Integrity

For each participant, scale scores for each of the Big Five factors and integrity were calculated for the two conditions separately. Table 1 presents the results. A within-participants multivariate analysis of variance (MANOVA) including all six scale scores showed that the means differed significantly between the two conditions, $F(6, 123) = 89.24, p < .001, \eta^2 = .81$. Subsequent paired-sample $t$ tests demonstrated that all scores were significantly higher in the faking than in the honest condition (with absolute $d$ values varying between 0.58 and 1.88), supporting the effectiveness of the fake good instructions. Consistent with our hypothesis, effect sizes indicated that participants inflated their scores the most on Conscientiousness and Emotional Stability. Scale scores for these two personality factors were well over 1.5 standard deviations higher in the fake good than in the honest condition, representing large effect sizes.

Response Latencies

Table 2 presents the means and standard deviations of the average response latency scores per Big Five factor and integrity in each condition. In total, across all items, response times were 0.25 s faster in the faking good than in the honest condition. This mean difference was significant, representing a small effect size ($d = 0.23$), and was robust for outliers. At the factor level, the differences in average response latencies between the two conditions were largest for Emotional Stability ($d = 0.41$) and Integrity ($d = 0.33$). No significant differences in average response latencies between the two conditions were found for Agreeableness and Autonomy.

Further, we tested Holden et al.’s (1992) interactive model of faking by comparing participants’ average response latencies on positively keyed and negatively keyed items. A $2 \times 2$ analysis of variance (ANOVA) with instruction (honest vs. fake good) and item desirability (positively keyed vs. negatively keyed) as within-participants factors demonstrated significant main effects for both instruction, $F(1, 128) = 7.89, p < .01, \eta^2 = .06$, and item desirability, $F(1, 128) = 16.03, p < .001, \eta^2 = .11$, with faking and positively keyed items leading to faster response latencies than answering honestly and negatively keyed items, respectively. However, no significant instruction by item desirability interaction effect was found.

Table 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Honest condition (ms)</th>
<th>Fake good condition (ms)</th>
<th>t</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Extraversion</td>
<td>3,403.43</td>
<td>1,604.62</td>
<td>3,124.26</td>
<td>1,058.94</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>3,201.06</td>
<td>1,303.25</td>
<td>3,125.85</td>
<td>1,087.10</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>3,248.71</td>
<td>1,060.30</td>
<td>3,013.64</td>
<td>1,032.14</td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>3,168.05</td>
<td>1,079.09</td>
<td>2,766.09</td>
<td>887.72</td>
</tr>
<tr>
<td>Autonomy</td>
<td>3,219.58</td>
<td>1,384.19</td>
<td>3,008.20</td>
<td>1,069.27</td>
</tr>
<tr>
<td>Integrity</td>
<td>4,216.52</td>
<td>1,695.59</td>
<td>3,682.23</td>
<td>1,528.15</td>
</tr>
<tr>
<td>Total</td>
<td>3,294.28</td>
<td>1,219.71</td>
<td>3,039.73</td>
<td>954.08</td>
</tr>
</tbody>
</table>

Note. $N = 129$.

*p < .05. **p < .01. ***p < .001.
was found, $F(1, 128) = 0.38, p = .54, \eta^2 = .00$. Thus, Holden et al.’s interactive model of faking was not supported.

In general, our findings on response latencies do not support the idea that faking takes more time than responding honestly, but rather that faking involves a faster response process for both positively and negatively keyed items.

**Eye Behavior**

**Total number of fixations.** For each participant, we first calculated the total number of eye fixations including all test items for the honest and faking condition separately. Table 3 (first data line) presents the mean and standard deviation across all participants of this total number of eye fixations for the two conditions. In total, participants had 98.92 fixations (i.e., slightly less than 1 fixation per test item on average) more in the honest condition ($M = 1,361.84$) than in the fake good condition ($M = 1,262.92$). This mean difference was significant, representing a small effect size ($d = 0.24$), and was robust for outliers. This finding corresponds with the results for response latencies, namely that faking is less cognitively demanding than answering honestly.

Further, we tested whether the number of fixations differed across the two conditions for positively keyed and negatively keyed items. For this purpose, we composed four variables for each participant, calculated as the total number of eye fixations per condition for positively keyed and negatively keyed items separately. Using these sum scores, a $2 \times 2$ with-in-participants ANOVA with instruction (honest vs. fake good) and item desirability (positively keyed vs. negatively keyed) as within-participants factors was conducted. Significant main effects were found for both instruction, and participants factors was conducted. Significant main effects were found across the two conditions for positively keyed and negatively keyed items separately. For this purpose, we composed four variables for each participant, calculating the total number of eye fixations across all test items) on each of the boxes of interest (i.e., q, a1–a5; see Figure 2) for each participant for the honest and faking condition separately. A within-participants MANOVA including all six boxes demonstrated that the number of fixations on the response options differed significantly between the two conditions, $F(6, 123) = 34.21, p < .001, \eta^2 = .63$. Table 3 presents the means and standard deviations across participants for these sum scores, as well as the subsequent $t$ tests. As hypothesized, when responding honestly, significantly more eye fixations occurred on the middle response options (a2, a3, and a4). In contrast, when faking, significantly more eye fixations occurred on the extreme response options (a1 and a5). The mean differences were significant, representing moderate to large effect sizes (with absolute $d$ values varying between 0.40 and 0.85), and were robust for outliers.

In addition, we examined the location of the fixations for positively keyed and negatively keyed test items separately. For this purpose, we composed four variables for each participant, calculated as the total number of eye fixations on box a1 per condition for positively and negatively keyed items separately. Using these sum scores, a $2 \times 2$ within-participants ANOVA was conducted, demonstrating that the response option a1 (much more [often] than others) was fixated on more for positively keyed items than for negatively keyed items, $F(1, 128) = 255.17, p < .001, \eta^2 = .67$. This difference between the number of eye fixations for positively and negatively keyed items was larger in the fake good than in the honest condition, as indicated by a significant interaction effect, $F(1, 128) = 49.85, p < .001, \eta^2 = .28$. Similarly, we composed the same four variables for each participant regarding the total number of eye fixations on box a5. The results of a $2 \times 2$ within-participants ANOVA demonstrated a reversed pattern for the eye fixations on response option a5 (much less [often] than others), such that the response option a5 was fixated on more for negatively keyed than for positively keyed items, $F(1, 128) =

### Table 3

**Means and Standard Deviations for the Average Number of Eye Fixations in the Honest and Fake Good Conditions, and Paired Samples $t$ Tests and Effect Sizes for the Differences Between the Two Conditions**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Honest condition</th>
<th>Fake good condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Total number of fixations</td>
<td>1,361.84</td>
<td>425.02</td>
</tr>
<tr>
<td>Number of fixations on:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question box (q)</td>
<td>671.67</td>
<td>194.17</td>
</tr>
<tr>
<td>Answer box 1 (a1)</td>
<td>62.07</td>
<td>44.81</td>
</tr>
<tr>
<td>Answer box 2 (a2)</td>
<td>176.84</td>
<td>89.92</td>
</tr>
<tr>
<td>Answer box 3 (a3)</td>
<td>134.40</td>
<td>67.90</td>
</tr>
<tr>
<td>Answer box 4 (a4)</td>
<td>106.41</td>
<td>54.51</td>
</tr>
<tr>
<td>Answer box 5 (a5)</td>
<td>25.02</td>
<td>21.76</td>
</tr>
</tbody>
</table>

Note. $N = 129$.

**p < .01. ***p < .001.**
302.73, \( p < .001, \eta^2 = .70 \). This difference between the number of eye fixations for positively and negatively keyed items was larger in the fake good than in the honest condition as indicated by a significant interaction, \( F(1, 128) = 91.81, p < .001, \eta^2 = .42 \).

Figure 3 displays the means for total number of eye fixations on response options a1 and a5 for positively and negatively keyed items in both conditions. Thus, especially in the faking good condition, people interpret the test items in terms of being positive or negative, and they consequently fixate on the much more (often) than others response option for positive items and on the much less (often) than others response option for negative items.

**Eye paths.** When looking to a test item screen, participants have various eye fixations spread across the screen. After each eye fixation, the eyes move to the next eye fixation (i.e., a saccade). In addition to the number and location of fixations, we analyzed the order of the fixations. For all participants, an eye path was composed for each test item describing the order of the fixation locations. For example, the path “q → q → a1 → a3 → a2” indicates that the participant’s first three fixations were on box q (i.e., at three different locations in box q), the fourth fixation was on box a1, the fifth fixation was on box a3, and the sixth fixation was on box a2. To facilitate data analysis and interpretation, repeated fixations on the same box of interest were combined. For example, the path mentioned above was transformed into “q → a1 → a3 → a2.” Analysis of all test items for all participants shows that 93.1% of the fixation paths start with a fixation on the question box (q), indicating that participants usually start with reading the question.

Next, we examined where the eyes move to directly after having read the question, by examining the first fixation after the fixation(s) on the question box. We were especially interested in eye paths indicating fixations on the extreme response options (a1 and a5) directly after having read the question (i.e., q → a1 and q → a5 eye paths), because such paths may be indicative of purely semantic item interpretations. For each participant, we calculated on how many positively keyed test items they had an eye path starting with a q fixation directly followed by an a1 fixation (i.e., q → a1 eye path), and on how many negatively keyed test items they had an eye path starting with a q fixation directly followed by an a5 fixation (i.e., q → a5 eye path) in the fake good and honest condition separately. In the fake good condition, participants had significantly more q → a1 eye paths on positively keyed items than in the honest condition (\( M = 14.41 \) vs. \( M = 9.07 \), \( t(128) = 7.93, p < .001, d = 0.67 \), and significantly more q → a5 paths on negatively keyed items (\( M = 3.02 \) vs. \( M = 1.09 \), \( t(128) = 5.38, p < .001, d = 0.59 \). These findings demonstrate that when in the faking good condition, after having read the question the eyes more often move directly to the extreme response option that fits with the item framing.

**Differentiating Faking From Honest Responding**

Based on the results of the analyses on actual responses (i.e., means), response latencies, eye fixations, and eye paths, we examined the extent to which fakers can be differentiated from honest responders. For this purpose, we analyzed our data between-participants (by taking the data of the first session only), because a between-participants design more accurately reflects actual settings in which one seeks to differentiate between fakers and honest responders. As such, we pooled the data from participants who were instructed to respond honestly in the first session (\( n = 64 \)) with the data from participants who were instructed to fake good in the first session (\( n = 65 \)). For each participant, we computed the following variables that may potentially add to the identification of faking. Regarding actual responses, for each participant we calculated the proportion of extreme responses on the test items. Regarding response latencies, based on the results as shown in Table 2, for each participant we calculated the mean response times on the Emotional Stability items, the Integrity items, the Conscientiousness items, and the Extraversion items. Regarding eye fixations, based on the results as displayed in Table 3 and Figure 4, for

![Figure 3](image.png)

*Figure 3.* Means for total number of eye fixations (on box d) for positively and negatively keyed items in the honest and fake good conditions.

![Figure 4](image.png)

*Figure 4.* Means for total number of eye fixations on response option boxes a1 (much more [often] than others) and a5 (much less [often] than others) for positively and negatively keyed items in the honest and fake good conditions.
each participant we calculated the mean number of fixations in box a1 for positively keyed items; the mean number of fixations in box a5 for negatively keyed items; and the mean number of fixations in boxes a2, a3, and a4 for all items. Regarding the eye paths, for each participant we calculated the average number of “q → a1” paths on positively keyed items, and the average number of “q → a5” paths on negatively keyed items.

Using these variables as predictors, we ran a hierarchical logistic regression analysis with response instruction on the first session (honest vs. fake good) as the dependent variable, and we computed the zero-order correlations (see Table 4 and Table 5). As shown in Table 4, the proportion of extreme responses and the response latency variables significantly differentiated fakers from honest responders, with a 79.8% classification hit rate. The eye fixation variables contributed significantly to the prediction over and beyond the extreme response and response latency variables, increasing the classification hit rate to 82.9% (with 83.1% of the fakers correctly identified as fakers, 16.9% of the fakers incorrectly identified as honest responders, and 17.2% of the honest responders incorrectly identified as fakers). The eye path variables did not further increase model fit.

Because half of our sample consisted of “fakers,” whereas in field settings the proportion of fakers is estimated to vary between 20% and 50% (Donovan et al., 2003; Griffith et al., 2007), we explored how our classification hit rate would hold under different conditions. Using R (R Development Core Team, 2010), we calculated 95% bootstrap confidence intervals (CIs) around the classification hit rate (using the percentile method; Efron, 1987), based on 1,000 bootstrap replications. Assuming a 50–50 distribution of fakers versus honest responders in a sample of 129, the 95% bootstrap CI was .783–.930, with a mean of .862. Assuming a 20–80 distribution in a sample of 129, the 95% bootstrap CI was .876–.977, with a mean of .929.

### Discussion

Both in science and in practice, faking on self-report personality measures is a much-debated issue. Research has demonstrated that individuals can and do fake on self-report non-cognitive measures, and that faking likely impacts both the construct and criterion-related validity. Because social desirability, impression management, or lie scales may not be very accurate to identify fakers, techniques that do not rely on self-report measures may be more useful for detecting faking behavior. In this context, previous research has examined people’s response latencies when faking (e.g., Holden, 1998; Holden & Hibbs, 1995; Vasilopoulos et al., 2000). The present study extends this line of research by not only examining response latencies but also people’s eye fixations and eye movements when responding to a Big Five measure and several Integrity items. As such, the present study adds to the literature by investigating whether eye-tracking can increase our understanding of people’s response processes when faking, and whether eye-tracking may yield information that can be used to identify faking behavior over and beyond the actual item responses and response latency information.

### Major Findings

Consistent with previous research (Alliger & Dwight, 2000; Viswesvaran & Ones, 1999), our findings demonstrate that people

---

**Table 4**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
<th>Step 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1: Response extremity index</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of extreme responses</td>
<td>1,460.03***</td>
<td>2,440.55***</td>
<td>8.8589</td>
<td>23.5857</td>
</tr>
<tr>
<td>Step 2: Response latencies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response time on Emotional Stability items</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9987*</td>
<td>0.9986†</td>
<td>0.9985†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response time on Integrity items</td>
<td>0.9996</td>
<td>0.9996</td>
<td>0.9997</td>
<td></td>
</tr>
<tr>
<td>Response time on Conscientiousness items</td>
<td>1.0013**</td>
<td>1.0019**</td>
<td>1.0019**</td>
<td></td>
</tr>
<tr>
<td>Response time on Extraversion items</td>
<td>1.0000</td>
<td>0.9997</td>
<td>0.9996</td>
<td></td>
</tr>
<tr>
<td>Step 3: Eye fixations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of a1 fixations on positively keyed items</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0135</td>
<td>1.0306</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of a5 fixations on negatively keyed items</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0381†</td>
<td>1.0230</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of a2 fixations</td>
<td>0.9858†</td>
<td>0.9852†</td>
<td>0.9852†</td>
<td></td>
</tr>
<tr>
<td>Number of a3 fixations</td>
<td>0.9859†</td>
<td>0.9843†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of a4 fixations</td>
<td>1.0180</td>
<td>1.0184</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 4: Eye paths</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of q → a1 paths on positively keyed items</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9016</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of q → a5 paths on negatively keyed items</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1793</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔChi² (df)</td>
<td>54.62 (1)***</td>
<td>15.04 (4)**</td>
<td>16.54 (5)**</td>
<td>2.87 (2)</td>
</tr>
<tr>
<td>Chi² (df)</td>
<td>69.66 (5)***</td>
<td>86.20 (10)***</td>
<td>89.06 (12)***</td>
<td></td>
</tr>
<tr>
<td>Nagelkerke R²</td>
<td>.46</td>
<td>.56</td>
<td>.65</td>
<td>.67</td>
</tr>
<tr>
<td>% correct classifications</td>
<td>77.5</td>
<td>79.8</td>
<td>82.9</td>
<td>82.9</td>
</tr>
</tbody>
</table>

**Note.** N = 129, with n = 64 in the honest condition and n = 65 in the fake condition.

*a* 0 = honest, 1 = fake good.

*p < .10,*  †*p < .05,*  ‡*p < .01,*  ‡‡*p < .001.
are able to fake on personality and integrity measures when instructed to do so. Means were 0.58–1.88 standard deviations higher in the fake good than in the honest condition, representing medium to large effect sizes. As hypothesized, differences between the conditions were largest for Conscientiousness and Emotional Stability (d = 1.88 and 1.69, respectively). This suggests that participants thought that high scores on these factors are most desirable, and that these factors are easiest to fake, most likely because it is relatively clear what a “good” response is for Conscientiousness and Emotional Stability items. Score differences between the two conditions were smallest for Agreeableness. Altogether, these findings demonstrate that participants have a relatively accurate idea of how to come across as a good applicant, because Conscientiousness and Emotional Stability have also been shown to be the Big Five factors with the strongest associations with job performance (Barrick et al., 2001).

The results on response latencies demonstrated that faking is on average 0.25 s faster than answering honestly, representing a small effect size (d = 0.23). At the personality factor level, the differences in response time between the two conditions were largest for Emotional Stability and Integrity items. For Emotional Stability, the difference between the two conditions can be explained by the low mean response time in the faking condition. More specifically, when faking good, people were fastest on the Emotional Stability items (about 0.30 s faster than on the other Big Five personality items). This finding further suggests that Emotional Stability items are relatively easy to fake. For Integrity, the large difference between the two conditions can be explained by the relatively high mean response time in the honest condition. Answering honestly on Integrity items takes on average 0.97 s more compared to answering honestly on the Big Five items. However, also faking good on Integrity items takes on average longer than faking good on the other items (about 0.64 s longer). The higher response latencies on Integrity items across the conditions may be caused by item length, because Integrity items were on average longer than the Big Five items (8.8 words compared to 5.6 words) and because item length is known to affect response latencies (Casey & Tryon, 2001; Holden et al., 1991). The finding that the difference between the Integrity and Big Five items was larger in the honest condition than in the fake good condition suggests that it is somewhat more difficult to answer honestly on Integrity items than on other personality items. Explanations may relate to the fact that the Integrity items cover more extreme aspects (e.g., stealing) or induce difficult thought-provoking situations (e.g., it may take some time to decide what to respond to an item such as “I am an honest person” when instructed to respond honestly).

The present study extends previous research by comparing people’s eye fixations when faking versus responding honestly. The results demonstrate that faking is associated with on average almost 1 eye fixation less per item compared to responding honestly. This finding is consistent with our results for response latencies and aligns with previous studies reporting lower response latencies when faking good (Hsu et al., 1989) and for items higher on social desirability (Holden et al., 1985). Furthermore, the eye fixation data demonstrated that participants paid more attention to the extreme response options in the fake good than in the honest condition. This finding was qualified by the item framing, such that participants paid more attention to the much more (often) than others response option for positively keyed items, and they paid

### Table 5: Between-Participants Correlations Among Proportion of Extreme Responses, Response Latencies, Eye Fixation Data, and Eye Path Data Using the First Test Session

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Proportion extreme responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Response time on Extraversion items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Response time on Agreeableness items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Response time on Emotional Stability items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Response time on Conscientiousness items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Number of a1 fixations on positively keyed items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Number of a2 fixations on positively keyed items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Number of a3 fixations on positively keyed items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Number of a4 fixations on positively keyed items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Number of q &lt;i&gt;a1&lt;/i&gt; paths on positively keyed items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Number of q &lt;i&gt;a2&lt;/i&gt; paths on positively keyed items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Number of q &lt;i&gt;a3&lt;/i&gt; paths on positively keyed items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Number of q &lt;i&gt;a4&lt;/i&gt; paths on positively keyed items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Number of q &lt;i&gt;a5&lt;/i&gt; paths on positively keyed items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 129, with n = 64 in the honest condition and n = 65 in the fake condition.
more attention to the much less (often) than others response option for negatively keyed items. In addition, eye path analyses showed that when faking good, people’s eyes more often moved directly to the extreme response option that corresponds to the item framing after having focused on the question text.

The findings on eye fixations were shown to be potentially useful for identifying faking behavior. More specifically, logistic regression analyses demonstrated that eye-tracking data improved correct classification of fakers beyond response extremity and response latency data. The response extremity data were able to distinguish fakers from honest-responders with a hit rate of 77.5%. Adding response latency data resulted in a 2.3% increment. Including data on the number of eye fixations on extreme and middle response options leads to a further improvement in distinguishing fakers from honest responders of 3.1%, up to a hit rate of 82.9% (which increased to an average of 93.0% in the bootstrap analyses). Eye path data, however, did not add over and beyond the eye fixation data, probably because of the relatively high intercorrelations between the response extremity, eye fixation, and eye path metrics (see Table 5). Nevertheless, the logistic regression results suggest that eye-tracking may be a useful addition for identifying faking behavior.

Theoretical Implications

As outlined in the Introduction, several contrasting theories have been proposed regarding the response processes when faking (e.g., Holden et al., 2001; Holtgraves, 2004; Vasilopoulos et al., 2000). Our results support the theoretical position that, compared to honest responding, faking is characterized by the less cognitively complex response process of direct retrieval. That is, when faking, respondents do not try to retrieve accurate information but rather produce a response solely based on the fake instruction combined with the social desirability framing of the item. In other words, faking more often leads to purely semantic rather than self-referenced item interpretations. More specifically, replicating previous research (Holden et al., 1985; Hsu et al., 1989), our response latency data show shorter response times when faking, indicating lower cognitive load. This lower cognitive load may result from differences in item interpretation between the two conditions, such that honest responding leads to self-referenced interpretations of the item content, whereas faking leads to direct retrieval or purely semantic item interpretations (i.e., does this item reflect something positive or negative?). Because semantic item interpretations take less processing time than self-referenced item interpretations (Rogers, Kuiper, & Kirker, 1977), our and previous response latency findings indirectly suggest that semantic item interpretations are more likely when faking.

Extending previous research, however, our eye-tracking data show more direct evidence for a semantic item interpretation characterization of the response process when faking good. That is, our findings on eye fixation locations and eye paths suggest that compared to honest responding, the response pattern when faking good more often is such that after having read the item, people directly fixate on the extreme response option that aligns with the item framing. Such eye paths are consistent with a direct retrieval or semantic item interpretation description of the response process when faking, because respondents apparently assessed the item content in terms of desirable or undesirable when reading the question, such that they could move immediately to the corresponding socially desirable extreme response option. In contrast, the pattern of results for answering honestly (i.e., slower response times, more fixations in general, more fixations on the middle response categories, and less eye paths from question directly to extreme response option) suggests that participants, when responding honestly, engage less in semantic item interpretations and make more considerate and self-referenced choices.

Our findings indicate that faking good leads to the adoption of an extreme response set (i.e., a disproportionate favor for the endpoints or extreme categories of the response scale; Lau, 2007; Naemi, Beal, & Payne, 2009). Previous research demonstrated that favoring extreme responses is related to individual differences such as intolerance of ambiguity, decisiveness, and simplistic thinking (Naemi et al., 2009); ethnicity (Bachman & O’Malley, 1984); and cultural differences such as power distance and masculinity (Johnson, Kulesa, Cho, & Shavitt, 2005). In addition to characteristics of the participant, several studies examined contextual factors (e.g., item format) that may induce adoption of an extreme response set (see Lau, 2007). Our findings suggest another important contextual factor that induces extreme responding, that is, a faking instruction or a high-stakes situation that evokes faking behavior. Future research is needed to determine the similarities and differences between extreme response styles (ERSs), social desirability, and faking. Lau (2007) posed that whereas ERS processes occur during the judgment phase in the response process, social desirability processes occur during the editing phase. Similarly, Holtgraves (2004) concluded that social desirability leads to the operation of a response editing mechanism. As demonstrated by the present study, deliberate faking, in contrast, occurs more immediately during the item interpretation and retrieval phases.

Combining these notions, we propose that faking refers to response processes characterized by semantic item interpretations and direct retrieval; ERS refers to response processes characterized by self-referenced item interpretations combined with a more extreme self-view during the retrieval and judgment phases, and social desirability refers to self-referenced item interpretations and judgments that are checked for social desirability during an additional response editing phase. Future research can apply eye-tracking technology to verify these proposed differences by comparing eye movements (e.g., eye fixations, eye paths) between a faking condition (e.g., using an instruction similar to the one in the present study) and a heightened social desirability condition (e.g., instructing that participants’ scores are used to create a personal profile; Holtgraves, 2004).

No support was found for Holden et al.’s (1992) interactive model of faking. For both response latencies and eye fixations, in addition to the main effect for instruction (honest vs. fake good) as discussed above, a main effect was found for item framing (positively vs. negatively keyed), such that positively keyed items were responded to faster and with less eye fixations. This finding corresponds with and extends previous research (Casey & Tryon, 2001), showing longer response latencies for negatively worded items than for positively worded items. No interactions indicating that the patterns for honest responding on positively and negatively keyed items were reversed from those for faking good were present. Although these findings seem to contradict Holden et al.’s interactive model of faking, previous studies usually found support for the model only for fake good versus fake bad comparisons.
rather than for fake good versus honest comparisons (e.g., Brunetti et al., 1998; Holden et al., 1992). Furthermore, previous research on the interactive model is mostly based on personality items with a dichotomous true–false response scale. By using such a scale it is clear which alternative is the socially desirable response and which alternative is the socially undesirable response. For items with a 5-point response scale as used in the present study, the social desirability of the response alternatives is more ambiguous. For example, research by Kuncel and Tellegen (2009) indicates that the most socially desirable response is not always the extreme response option that corresponds with the item framing. Thus, based on these findings, future research should investigate whether Holden et al.’s interactive model of faking may hold for items with 5-point Likert scales when adjusting for response option desirability, as demonstrated by Kuncel and Tellegen.

One remarkable inconsistency in the present study findings is that, on the one hand, the response latency and eye-tracking data clearly suggest that faking is cognitively less complex and easier than responding honestly, whereas, on the other hand, the results on the perceived task difficulty items suggest that participants found faking significantly more difficult than answering honestly. Although this finding remains intriguing, a possible explanation may be that participants felt that faking was not so much difficult in the sense of being cognitively complex but more in terms of being ethically difficult. Another possible explanation might be social desirability responding on the task difficulty items, such that the participants thought that it would not be good to find faking very easy.

Limitations and Boundary Conditions

Several boundary conditions and limitations have to be taken into account when interpreting the present findings. An important boundary condition is the study setting, using instructed faking. Because people were instructed to fake, they probably did not have any concern with the accuracy of their responses and needed not to fear that their faking behavior was detected and/or punished. This situation likely differs from actual selection situations, in which participants may have more concerns with the accuracy of their responses (thus increasing self-referenced interpretations) and may fear that faking behavior is detected and has negative consequences. Such situations could cause people to engage in less extreme and more considerate faking behavior, characterized by more complex response processes, resulting in higher cognitive load, slower processing, and probably more eye fixations. These aspects may explain the differences between the present results and some previous studies. For example, our findings might be viewed as inconsistent with the literature on lying and deceiving (e.g., DePaulo et al., 2003), which suggests that lying is cognitively more complex (Vrij et al., 2001; Zuckerman et al., 1981). However, in the present study, participants were not so much asked to lie but to respond to the test items such that they came across as the ideal employee. Also, our response latency findings differed from those as reported by McDaniel and Timm (1990) and Holtgraves (2004), which probably can be explained by the type of manipulation used to induce faking. Whereas in the present study participants were asked to respond such that they would come across as the most suitable job applicant (cf. McFarland & Ryan, 2000; Vasilopoulos et al., 2000; Zickar & Robie, 1999), McDaniel and Timm asked participants to answer the questions exactly the same way that they would if they were taking the test for a job they were interested in obtaining, and Holtgraves instructed participants that their scores would be used to create a personal profile. These latter instructions both include elements that likely induce more self-referenced item interpretations and response editing mechanisms, which may explain the longer response times. Although it can be debated which instruction is best, it is probably likely that both types of response processes (i.e., faking in terms of purely semantic item interpretations vs. social desirability responding in terms of more self-references item interpretations combined with a response editing mechanism) will occur in test settings. Consistent with this idea, Zickar, Gibby, and Robie (2004) found that among actual job applicants, one can distinguish between a slight faking group and an extreme faking group, in addition to honest respondents. Future research should therefore distinguish between several degrees/types of response distortion and compare the effects of different faking or social desirability instructions on response latencies and eye behavior.

With the exception of the logistic regression analyses for distinguishing between fakers and honest-responders, we analyzed our data using a within-participants design. Although such a design is the most appropriate for understanding differences in response processes between faking and honest-responding, because it removes possible effects of individual differences in response latencies and eye behavior, it may be argued to reduce the generalizability of the findings to operational settings. Therefore, we repeated all our analyses in a between-participants design, using the data from the first test administration only. These analyses demonstrated similar patterns regarding the differences between the two conditions, illustrating the robustness of our findings.

Before the test sessions started, participants were told that their eye movements were tracked, which could be argued to have affected our findings. Although we cannot completely rule out this possibility, we think it is unlikely. That is, participants completed 105 test items under a fake good and under an honest instruction. Participants were not aware of the study purpose and hypotheses or of the reasons for the eye-tracking. Although people can control their eye movements to some extent, it takes considerable self-regulatory effort (Everling & Fisher, 1998). For these reasons, we regard it unlikely that our results are caused by participants consciously manipulating their eye movements differentially between the two conditions and consistently over 105 test items.

Future research is needed to cross-validate and examine the robustness of our findings in actual selection settings. For example, research should examine the generalizability of our findings to different samples, different personality measures, and different response scales (e.g., ipsative tests, true–false items). Although we could have cross-validated our logistic regression findings using the second test administration, future research with a fresh sample of participants who are not influenced by a previous test session is deemed preferable. Regarding the robustness of our findings in actual selection settings, as we explicitly instructed the participants to fake good, the response processes may have been different from the response processes of fakers in actual selection contexts. Participants may have been less tense, as there was no threat of being caught on faking. Future research on eye movements and faking should study differences in eye movement between job applicants...
and job incumbents. Also, manipulations may be used that better mimic actual faking conditions, for example, by warning participants in the faking condition that they should try to fake without getting caught.

In the present study, test items were categorized into positively and negatively keyed items, because item framing affects response latencies and eye behavior. Recent work by Kuncel and Tellegen (2009) indicates that items can be further classified based on the socially desirable profile of the response options. Whereas some items (e.g., “I am punctual”) showed a linear profile indicating that more endorsement is more socially desirable, other items (e.g., “I am talkative”) were found to show inverted U-shape patterns indicating that the middle response options were most socially desirable. An interesting avenue for future research relates to examining whether such item characteristics may lead to more specific insight in the response processes when faking by studying people’s eye behavior per item type.

Lastly, future research should investigate moderators in the relationship of faking/answering honestly with response latencies and eye behavior. For example, previous faking research demonstrated that Integrity, Conscientiousness, and Emotional Stability are negatively related to faking behavior (McFarland & Ryan, 2000) and that job familiarity affects how impression management relates to response latencies (Vasilopoulos et al., 2000). Extending these studies, future research could investigate to what extent cognitive ability, personality, job familiarity, knowledge of and/or previous experience with personality tests, emotional arousal, and item content influence the effects of faking on response latencies and eye behavior.

Practical Implications and Conclusion

As far as we know, this study is the first study on eye-tracking and faking on personality and integrity measures. The eye-tracking and response latency findings were supportive of the idea that instructed faking on personality tests is characterized by a faster and less cognitively demanding response pattern, characterized by semantic item interpretations and direct retrieval. Furthermore, differences in eye behavior between honest responding and faking good were found to be useful in identifying faking behavior. For example, having many fixations on extreme response categories (directly after having read the question) and fewer fixations on middle response categories may be signs of faking. Although future research is needed to replicate these results, based on the present findings, it may be useful to develop faking indicators based on eye-tracking measures. However, the development of such indicators should be based on tailor-made research for the particular personality test that is used. Furthermore, in addition to global indicators across groups of test items (as we used), it may be valuable to examine whether identification of faking can be improved by developing indicators based on specific items that differentiate strongly between fakers and honest responders.

A possible factor that may limit the practical applicability of eye-tracking is the eye-tracking equipment. Although the technology that we used is much less intrusive than older eye-tracking devices, it may still be considered impractical. Other eye-tracking devices are available that use a laptop computer with a built-in eye-tracking camera, which may be relatively easy to use in operational settings. However, because of calibration, respondents will always be aware of the fact that their eye behavior is recorded. Thus, similar to using response latencies or social desirability scales to detect faking, test-takers may potentially be able to influence their eye movements, although it would demand considerable test knowledge and self-regulation to consistently influence one’s own eye movements over an entire test session. Nevertheless, this is an important issue that has to be examined in greater detail before applying eye-tracking in practice.

In conclusion, despite potential limitations regarding the external validity, the present study contributes to the literature on faking by demonstrating differences in eye behavior when faking good on personality test items compared to answering honestly, using innovative eye-tracking technology. Eye-tracking may be promising for the field of personality and personnel selection, as it was demonstrated to be useful in identifying faking behavior over and beyond using response extremity and latency data. In addition to detecting faking behavior, eye-tracking may be an interesting tool for developing non-self-report personality tests for traits such as optimism or anxiety, as previous research demonstrated that eye movements when looking to visual stimuli relate to trait optimism (Isaacowitz, 2005) and trait anxiety (Calvo & Avero, 2005).

References


Received March 23, 2009
Revision received August 8, 2011
Accepted August 26, 2011