Judgment and Decision Making, Vol. 13, No. 4, July 2018, pp. 356–371

Who lies? A large-scale reanalysis linking basic personality traits to unethical decision making

Daniel W. Heck* Isabel Thielmann† Morten Moshagen‡ Benjamin E. Hilbig†

Abstract

Previous research has established that higher levels of trait Honesty-Humility (HH) are associated with less dishonest behavior in cheating paradigms. However, only imprecise effect size estimates of this HH-cheating link are available. Moreover, evidence is inconclusive on whether other basic personality traits from the HEXACO or Big Five models are associated with unethical decision making and whether such effects have incremental validity beyond HH. We address these issues in a highly powered reanalysis of 16 studies assessing dishonest behavior in an incentivized, one-shot cheating paradigm (N = 5,002). For this purpose, we rely on a newly developed logistic regression approach for the analysis of nested data in cheating paradigms. We also test theoretically derived interactions of HH with other basic personality traits (i.e., Emotionality and Conscientiousness) and situational factors (i.e., the baseline probability of observing a favorable outcome) as well as the incremental validity of HH over demographic characteristics. The results show a medium to large effect of HH (odds ratio = 0.53), which was independent of other personality, situational, or demographic variables. Only one other trait (Big Five Agreeableness) was associated with unethical decision making, although it failed to show any incremental validity beyond HH.

Keywords: cheating; dishonesty; logistic regression; HEXACO Honesty-Humility; Big Five

1 Introduction

Dishonest behavior is prevalent in various life settings. On the societal level, dishonesty incurs substantial costs and thereby poses a major threat to society at large (Mazar & Ariely, 2006). Given the importance of unethical decisions for societal functioning, research across various disciplines has addressed which situational factors and personality traits can account for dishonesty, both in isolation as well as in interaction with each other.

A common approach to study unethical decision making are so-called cheating paradigms from behavioral ethics, such as the coin-toss task (Bucciol & Piovesan, 2011) or the dice-roll task (Fischbacher & Föllmi-Heusi, 2013). In the coin-toss task, participants toss a coin in private and are asked to report the outcome. They know that one side of the coin (e.g., Heads) is associated with a monetary payoff whereas the other (e.g., Tails) is not. In turn, given that the coin is tossed in private, the actual outcome is unknown to the experimenter and participants who actually got Tails can easily cheat by misreporting when having obtained Heads to receive the associated payoff. Similarly, in (a variant of) the dice-roll task, participants are asked to roll a fair die in private and to report whether they obtained a specific "target number" (e.g., a four) which is associated with a monetary payoff. If they respond "yes", they receive the payoff; if they respond "no", they receive nothing. Again, since the die is rolled in private, participants can simply respond "yes" irrespective of their actual outcome to receive the payoff. Importantly, because one can honestly obtain the favorable outcome by luck, the response associated with the favorable outcome is not self-incriminating (Moshagen, Hilbig, Erdfelder & Moritz, 2014). However, given that the baseline probability p of observing a favorable outcome is known by design, (i.e., $p = 1/2$ in a single coin toss with a fair coin and $p = 1/6$ in a single dice roll with a fair, six-sided die), the proportion d of dishonest individuals who are prepared to lie can be estimated at the aggregate level (Moshagen & Hilbig, 2017).[1]

Data and R code are available at the Open Science Framework: https://osf.io/56hw4.

DWH was supported by the research-training group *Statistical Modeling in Psychology*, funded by the German Research Foundation (GRK 2277).

BEH and MM were supported by grants from the German Research Foundation (HI 1600/1-2 and MO 2830/1-1, respectively).

We thank Emmanuelle Kleinlogel, Sascha Müller, Marco Perugini, and Stefan Pfattheicher for providing their datasets for this reanalysis and the permission to publish the data online.

*Statistical Modeling in Psychology, University of Mannheim, B6, 30–32 (room 313), 68159 Mannheim, Germany. E-Mail: heck@uni-mannheim.de.

†University of Koblenz-Landau

‡Ulm University

[1]In contrast to the proportion of dishonest individuals d, the proportion of *cheaters* c refers to participants who actually lied (i.e., observed the unfavorable outcome and responded "yes"), that is, $c = (1 - p)d$.

1.1 Individual Differences in Dishonesty

A key finding from studies using cheating paradigms as described above is that some – but not all – individuals are dishonest. That is, the empirical proportion of "yes"-responses usually exceeds the baseline probability p while still being considerably smaller than one (Abeler, Nosenzo & Raymond, in press). Strikingly, the proportion of dishonest individuals remains relatively stable, even when stakes are raised substantially (Hilbig & Thielmann, 2017; Kajackaite & Gneezy, 2017). Indeed, it appears that most individuals are willing to cheat only a little (Hilbig & Hessler, 2013; Shalvi, Handgraaf & De Dreu, 2011) – arguably to protect their positive self-image, even in the face of dishonesty (Dana, Weber & Kuang, 2007; Mazar, Amir & Ariely, 2008; Shalvi et al., 2011). However, some individuals entirely refrain from lying and others actually lie to the maximum extent possible (e.g., Fischbacher & Föllmi-Heusi, 2013; Hilbig & Thielmann, 2017). Thus, the empirical picture consistently shows that individuals strongly differ in their willingness to lie, which raises the question how to account for this observed heterogeneity in individual behavior.

Previous research has often focused on personality to account for the apparent individual differences in unethical decision making. Most consistent evidence on the link between (basic) personality traits and dishonesty has been accumulated for the Honesty-Humility (HH) dimension of the HEXACO model of personality (Ashton & Lee, 2007), which incorporates characteristics such as Sincerity, Fairness, Modesty, and Greed Avoidance. Corresponding to this conceptualization, studies have shown a substantial negative association between HH and dishonest behavior in cheating paradigms (e.g., Hilbig & Zettler, 2015; Kleinlogel, Dietz & Antonakis, 2018; Thielmann, Hilbig, Zettler & Moshagen, 2017). For other basic traits within and beyond the HEXACO taxonomy, in turn, no consistent pattern has been observed (e.g., Hilbig, Moshagen & Zettler, 2016; Hilbig & Zettler, 2015; Williams, Nathanson & Paulhus, 2010).

However, although the negative link between HH and dishonesty has been repeatedly and consistently shown, several open questions remain. First, there is no precise effect size estimate available so far, given that studies investigating said link had comparably small sample sizes (median N = 196). This issue is further aggravated by the low efficiency (and thus, statistical power) of studies using cheating paradigms, because individuals' responses are associated with a larger measurement error due to the random noise added by the chance mechanism (coin toss or die roll). This random noise, in turn, has often been neglected in analyses using standard logistic regression to model the link between personality traits and "yes"-responses (e.g., Hilbig & Zettler, 2015; Zettler, Hilbig, Moshagen & de Vries, 2015). Thus, previously reported effect size estimates were systematically downward biased (Moshagen & Hilbig, 2017; see below). However, given that cumulative science needs to go beyond testing *whether* a link exists, but rather provide an unbiased and precise estimate of the corresponding effect size (Cumming, 2014; Meehl, 1990), the current state of knowledge remains unsatisfactory.

Second, it is currently unclear whether and to what extent other basic personality traits aside from HH can account for dishonest behavior. Although prior studies did not unravel consistent associations, other basic traits beyond HH might actually predict dishonesty, albeit to a smaller degree. For instance, some studies have revealed significant effects of Big Five Agreeableness and both Big Five and HEXACO Conscientiousness on dishonesty (Hilbig & Zettler, 2015; Horn, Nelson & Brannick, 2004; Williams et al., 2010) whereas others found no links at all (Jones & Paulhus, 2017), or only for some Big Five inventories (Hilbig et al., 2016). Partly, such inconsistencies might be attributable to limited statistical power of studies, which makes it difficult – if not impossible – to distinguish between existing but small effects and null effects. Indeed, a sensitivity simulation (see Appendix) shows that very large sample sizes ($N \geq$ 3,200) are required to ensure a high probability ($\geq$ 72%) of obtaining moderate evidence for or against small effects of personality traits on dishonesty.

Third, it is currently unclear whether the HH-cheating link is at least partially attributable to covariance of HH with other variables, meaning that HH might be less predictive for dishonesty in the context of these other variables. In particular, the effect of HH may diminish once taking relevant demographic covariates such as sex and age into account. In a recent meta-analysis, Abeler et al. (2016) showed that women cheat less than men and also found a (nonsignificant) trend that older participants cheat less than younger participants (see also Bucciol, Landini & Piovesan, 2013). Given that women have somewhat higher levels of HH than men (Lee & Ashton, in press; Moshagen, Hilbig & Zettler, 2014) and that HH is positively correlated with age (Ashton & Lee, 2016), the HH-cheating link might thus be spurious and reflect a sex-cheating or an age-cheating link to some (unknown) degree.

Fourth, the limited statistical efficiency and power of prior single studies hampers investigation of trait interactions, which will arguably resemble only small effects. Indeed, the theoretical conceptualizations of the remaining HEXACO dimensions imply that other trait dimensions might moderate the HH-cheating link. A first candidate in this regard is Conscientiousness. As proposed by Lee and Ashton (2013, p. 59) "people who have both the exploitiveness of low H [Honesty-Humility] and the impulsiveness of low C [Conscientiousness] are doubly inclined toward criminal behaviour in general." In contrast, individuals low in HH and high in Conscientiousness should be "less inclined to break laws, even though they're perfectly willing to exploit other people, because they like rules and order" (Lee & Ashton, 2013, p.

59). Thus, in a cheating paradigm, those low in HH and low in Conscientiousness should be more likely to follow their first impulse to behave dishonestly whereas those low in HH and high in Conscientiousness may be better at controlling their first impulses and rather adhere to the (in this case ethical) rules. This prediction of a disordinal interaction of HH and Conscientiousness is in line with research showing that honesty requires attention to and consideration of normative ethical rules (Shalvi, Eldar & Bereby-Meyer, 2012).

Moreover, a priori hypotheses can be generated for Emotionality as another potential moderator of the HH-cheating link. Specifically, "people who combine low Honesty-Humility with low Emotionality have a whole lot of greed and not much fear" whereas "low-H, high-E people will try to exploit others, but they will do so in subtle, sneaky ways in order to avoid any confrontation or other risk of harm" (Lee & Ashton, 2013, pp. 40–42). Individuals low in HH but high in Emotionality might therefore be inclined to lie (low HH) but nonetheless refrain from doing so because they are afraid of negative consequences (high Emotionality). This may be the case although lying is fully concealed by design in the cheating paradigm because those high in Emotionality might transfer their generally high level of anxiety to this situation and therefore act *as if* they would need to fear sanctions by others. Moreover, they might fear the internal costs caused by lying due to threatening one's moral self-image (Mazar et al., 2008). Their counterparts low in both HH and Emotionality, in turn, should be less fearful of potential negative consequences of unethical decisions and therefore simply follow their inclination to lie. Overall, these predictions imply a disordinal interaction of HH and Emotionality on dishonest behavior.

Fifth, and finally, it is largely unknown whether the HH-cheating link is subject to any boundary conditions related to design factors of the cheating paradigm. Arguably, the most important characteristic of cheating paradigms is the baseline probability p of observing the favorable outcome which may differ across studies depending on the randomization device implemented (e.g., $p = 1/2$ in a single coin toss vs. $p = 1/6$ in a single die roll). In turn, it has recently been shown that the likelihood to behave dishonestly increases with a higher baseline probability p (Abeler et al., in press), a finding that is plausibly attributable to a general preference to appear honest (Utikal & Fischbacher, 2013). Since the chances of legitimately receiving a payoff are generally greater the larger the baseline probabilities p, "yes"-responses become less suspicious and thus less self-incriminating. This increase in anonymity might be especially relevant for individuals who have a general inclination to cheat (i.e., those low in HH): Because these individuals should be motivated to save their face as an honest person (Hilbig, Moshagen & Zettler, 2015), these individuals might particularly consider the probability with which they could, in principle, be exposed as a cheater. In contrast, those high in HH should refrain from lying irrespective of the baseline probability p. Thus, the baseline probability might moderate the HH-cheating link such that it should be larger the higher p.

To answer these five open questions related to the link between basic personality traits and unethical decision making, we reanalyze 16 studies on the level of the raw data with a total sample size of $N = 5{,}002$ using a newly developed, generalized regression model for one-shot cheating paradigms. Based on the large sample size, our analysis allows for a much higher statistical precision in estimating the effect size of the HH-cheating link and provides sufficient statistical power for testing small versus null effects of other traits and interactions of HH (see sensitivity simulation in the Appendix).

2 Methods

2.1 Datasets

Our reanalysis focuses on studies using cheating paradigms with real (monetary) incentives and dichotomous "yes"/"no"-responses (most prominently, the coin-toss and the dice-roll task) for the following reasons: First, using incentivized cheating paradigms allows drawing conclusions on *actual* behavior (Baumeister, Vohs & Funder, 2007) in the sense that decisions are linked to real consequences. Second, dishonesty in cheating paradigms – if implemented in a one-shot ("yes"/"no") fashion, as focused on in our reanalysis – is not self-incriminating given that "yes"-responses may also result from getting lucky and thus be legitimate. Third, the exact probability of legitimate "yes"-responses (i.e., the baseline probability p) is known by design. This allows adjusting for the additional, unsystematic source of random noise in "yes"-responses, thereby providing an unbiased estimate of the probability of dishonest behavior (see below).

To identify articles suited for our reanalysis, we searched the database *PsycINFO* for published studies that used a binary, incentivized cheating paradigm and also measured HH with a version of the HEXACO Personality Inventory-Revised (Lee & Ashton, 2004). Since the statistical model assumes that responses are dichotomous (see limitations in the Discussion), we did not include studies that implemented non-binary versions of the cheating paradigm. For instance, we excluded studies that asked participants to report the total number of favorable outcomes they obtained across multiple dice rolls (Kleinlogel et al., 2018). Additionally, we considered unpublished studies from our own and related labs. Overall, this resulted in a total number of 16 studies as summarized in Table 1.[2] Of these studies, all but one (Perug-

[2]Due to full anonymity of participants, personal identifiers were not available to match individuals across datasets. Hence, it is possible that the same participants are included in more than one dataset, thus entering the reanalysis twice. However, this is unlikely given that the three student

https://doi.org/10.1017/S1930297500009232 Published online by Cambridge University Press

Table 1: Datasets included in the reanalysis.

	Study	N	% Female	Age (SD)	Students	Delay	Baseline p	P("yes")	Incentive	NEO-FFI
Hilbig & Zettler (2015)	2	88	61.4	21.4 (3.8)	x	x	.167	.455	5€	
Hilbig & Zettler (2015)	5	147	53.1	39.9 (14.1)			.161; .322*	.418	2€; 4€*	x
Hilbig & Zettler (2015)	6	107	42.1	43.9 (13.2)		x	.25	.505	5€	x
Hilbig, Moshagen & Zettler (2016)		929	52.4	39.9 (13.3)			.375	.576	5€	x[†]
Klein, Thielmann, Hilbig & Zettler (2017)		210	49.0	40.0 (12.9)			.25	.440	5€	
Klein, Thielmann, Hilbig & Zettler (2018)		57	64.9	43.5 (12.2)		x	.25	.526	5€	x
Moshagen (2016)		650	57.9	47.0 (13.9)		x	.25	.263	5€	
Moshagen, Hilbig & Zettler (in press)	3	882	46.0	42.5 (13.0)		x	.25	.374	5€	x
Müller & Moshagen (2018)		460	48.3	47.3 (13.4)		x	.25	.359	5€	
Perugini & Leone (2009)	1	37	32.4	24.9 (2.9)	x		0[§]	.162	15 GBP[§]	
Pfattheicher & Schindler (in press)	1	192	49.5	22.5 (4.4)	x		.333	.609	5€	
Pfattheicher & Schindler (in press)	2	472	59.5	38.5 (12.0)			.50	.742	$0.25	
Thielmann, Hilbig, Zettler & Moshagen (2017)	2	152	51.3	51.5 (14.3)			.25	.500	5€	
Thielmann & Hilbig (in press)	1	183	49.2	37.6 (12.4)		x	.25	.503	5€	
Thielmann & Hilbig (2018)	2	200	48.0	43.8 (12.9)		x	.25	.450	5€	
Thielmann, Klein & Hilbig (2018)	1	89	42.7	42.3 (13.0)			.25	.506	5€	

* The probability p was a within-subject factor, whereas incentive was a between-subject factor.

[†] Only 310 participants completed the NEO-FFI.

[§] Perugini and Leone (2009) filmed the die roll and excluded participants from the analysis if the die showed the favorable outcome (i.e., a legitimate win). Moreover, the payoff was probabilistic, that is, a lottery in which participants could win 15GBP, and participants were unaware of the actual baseline probability p because the die was biased.

ini & Leone, 2009) measured all six HEXACO dimensions. Overall, the reliabilities of the six HEXACO scales, each composed of 10 items, were satisfactory (HH: $\alpha = .74$, Emotionality: $\alpha = .76$, Extraversion: $\alpha = .81$, Agreeableness: $\alpha = .72$, Conscientiousness: $\alpha = .75$, Openness to Experience: $\alpha = .78$). Moreover, five of the studies measured the Big Five using the NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992).[3] The total sample size across these studies amounted to $N = 5{,}002$ responses from 4,855 participants (i.e., 147 participants in Study 5 of Hilbig & Zettler, 2015, completed the cheating paradigm twice with varying baseline probabilities in a within-subject design).

Several of the included studies used (approximatively representative) community samples rather than student samples. Correspondingly, the average age of participants across studies was $M = 41.3$ ($SD = 14.1$) and participants were essentially equally distributed across the sexes (with 51.4% female participants). This considerable heterogeneity of the sample ensured high variability of personality traits – which is a precondition for identification of associations with external criteria. The merged dataset as well as all R scripts for the analysis are openly accessible at https://osf.io/56hw4.

samples were collected by different research groups at different universities and that the remaining datasets featured community samples in Germany recruited by different professional panel providers via the internet.

[3]Hilbig et al. (2016) compared three different Big Five inventories. However, we only include Big Five scores for those 310 of the 929 participants who completed the NEO-FFI to ensure that all Big Five indicators are based on the same inventory.

2.2 Statistical Analysis of the Cheating Paradigm

Our hypotheses concern the association between continuous variables (personality traits) and the probability of dishonest behavior in cheating paradigms (i.e., "yes"-responses). Since standard logistic regression provides biased effect size estimates, Moshagen and Hilbig (2017) proposed a modified logistic regression approach that accounts for the statistical structure of cheating paradigms. However, this method requires that the baseline probability p is constant for all observations and that all responses are independent and identically distributed. Since both of these assumptions do not hold in our reanalysis of the merged raw data of 16 studies, we developed a generalized version of the regression model.

Modified Logistic Regression. The modified logistic regression (Moshagen & Hilbig, 2017) directly models the probability of dishonesty d that participants respond "yes" irrespective of the actual outcome. For this purpose, the approach explicitly takes the baseline probability p of "yes"-responses into account. For a participant i, the probability of a "yes"-response is modeled as a sum of two terms: Either, the participant is prepared to lie with probability d_i and responds "yes" irrespective of the observed outcome; or she behaves honestly with probability $1 - d_i$ and responds "yes" due to observing the favorable outcome with probability p. Thus, the total probability that an individual responds "yes" is:

$$P(Y_i = \text{"yes"}) = d_i + (1 - d_i)p \tag{1}$$

Whereas the baseline probability p is determined by the design of the cheating paradigm and thus known a priori (e.g., $p = 1/2$ in a single coin toss), researchers are usually interested in estimating the probability of dishonesty d_i of an individual i. To test whether the probability of dishonesty is associated with covariates such as personality traits, the individual probabilities d_i are modeled in a logistic regression with design matrix $\mathbf{X}$ and a vector of regression weights $\boldsymbol{\beta}$ (Moshagen & Hilbig, 2017):

$$d_i = \frac{\exp(\mathbf{X_i}\boldsymbol{\beta})}{1 + \exp(\mathbf{X_i}\boldsymbol{\beta})} \tag{2}$$

where $\mathbf{X_i}$ is the i-th row of the design matrix containing the predictor values for the i-th individual. When plugging the logistic model in Equation 2 into the likelihood function of the data (Equation 1), one obtains the complete model for the probability of a "yes"-response. In our reanalysis, the design matrix $\mathbf{X}$ includes measures of the personality traits, situational covariates such as the baseline probability p, and multiplicative interaction terms. Maximum-likelihood estimates $\hat{\boldsymbol{\beta}}$, standard errors, and significance tests for the regression parameters are easily obtained using the R package RRreg (Heck & Moshagen, 2018).

Generalized Regression Framework for Cheating Paradigms. The modified logistic regression model by Moshagen and Hilbig (2017) requires that the baseline probability p is constant for all responses and participants. However, the studies included in our reanalysis implemented the cheating paradigm with different random devices and varying ps. We therefore generalize the existing approach by assuming a link function that varies depending on the baseline probability p_i for each observation:

$$P(Y_i = \text{"yes"}) = d_i + (1 - d_i)p_i. \tag{3}$$

The adjustment of the link function for each observation is necessary because for the same level of dishonesty, the expected proportion of "yes"-responses depends on the baseline probability p_i. For instance, the probability of "yes"-responses for individuals one standard deviation below the mean on a certain covariate (e.g., HH) is 11.1%, 25.9%, and 40.7% for baseline probabilities of $p = 0$, $p = 1/6$, and $p = 1/3$, respectively. If the model does not account for such different baseline probabilities explicitly, the statistical analysis will result in biased estimates of the regression coefficients.[4]

The modified logistic regression by Moshagen and Hilbig (2017) also assumes that observations are independent and identically distributed, which is not the case for the nested data in our reanalysis. As a remedy, we assume a hierarchical structure with "yes"-responses on level-1 that are nested within study on level-2. Technically, we add a random-intercept on the logit scale in Equation 2, which allows for different overall levels of dishonesty per study (i.e., varying means of the probability of dishonesty d_i. Similar as in hierarchical logistic regression or hierarchical multinomial models (e.g., Heck, Arnold & Arnold, 2018), the random effects are assumed to follow a centered normal distribution on the group level, with a standard deviation σ that quantifies the heterogeneity in dishonesty across studies.

We implemented the generalized model in a Bayesian framework (Wagenmakers, 2007), which has the advantage that credibility intervals can be intuitively interpreted as the most plausible range of the parameters given the data (Hoekstra, Morey, Rouder & Wagenmakers, 2014). Moreover, Bayes factors provide a continuous measure of the relative evidence for versus against an effect (Wagenmakers, 2007). For instance, a Bayes factor of $B_{10} = 3$ implies that the odds in favor of the alternative hypothesis H1 (assuming a nonzero effect) against the null hypothesis H0 (assuming a zero effect) increase by a factor of three after observing

[4]This issue is especially severe if the baseline probability p_i is correlated with any covariate in the design matrix. As an example, consider the scenario of a covariate that is independent of the probability of dishonest behavior but related to the baseline probability p_i (e.g., a dummy variable coding whether a die or coin served as random device). A standard logistic regression will indicate that this covariate has an effect on the probability of "yes"-responses – but spuriously so, because the larger proportion of "yes"-responses is falsely attributed to the covariate that correlates with the baseline probability p_i.

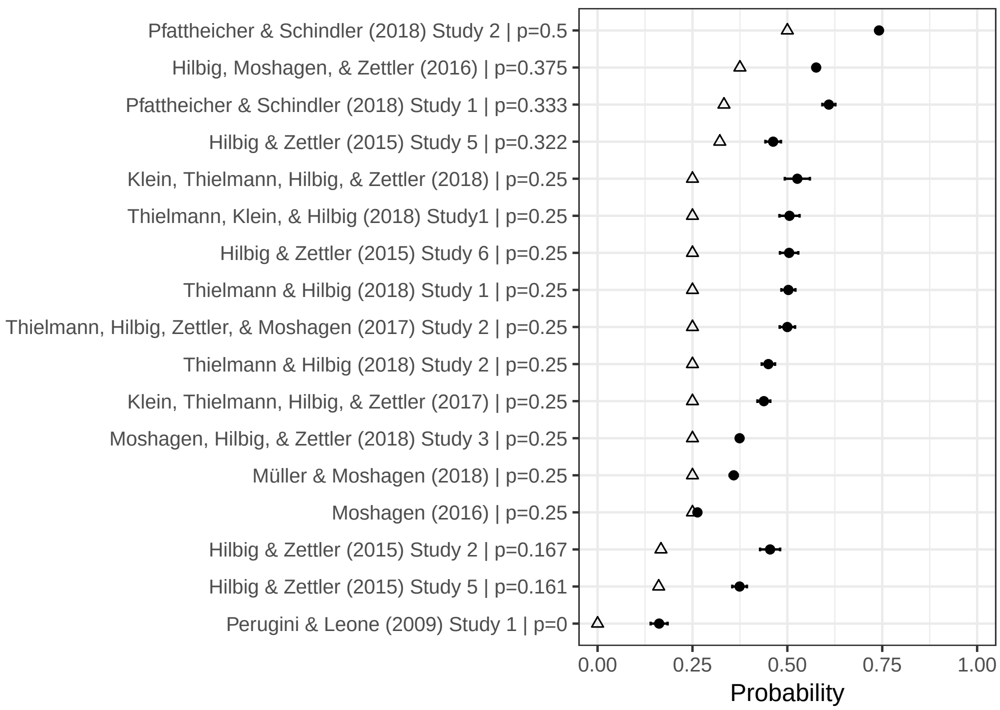

FIGURE 1: Baseline probabilities p (triangles) and proportion of "yes"-responses (solid points; error bars +1/−1 SE) in the studies included in the reanalysis.

the data. Assuming equal prior odds of 1:1 for the two hypotheses, this means that the posterior odds become 3:1 (i.e., the data provide evidence for a nonzero effect). Vice versa, the inverse $B_{01} = 1/B_{10}$ quantifies the evidence for the null hypothesis H0 relative to the alternative hypothesis H1. Bayesian inference requires prior distributions for the model parameters, which need to be chosen carefully. We adapted default priors for linear regression models (i.e., the Jeffreys-Zellner-Siow prior; Liang, Paulo, Molina, Clyde & Berger, 2008) that have desirable theoretical properties (e.g., scale invariance and consistency) and have become the standard in psychology for computing Bayes factors (Rouder & Morey, 2012).

The supplementary material (https://osf.io/56hw4) provides a more technical definition of the generalized modified logistic regression model and also the R code to fit the model with the software Stan (Carpenter et al., 2017). Bayes factors are computed with bridge sampling (Gronau, Wagenmakers, Heck & Matzke, 2017) and the generalized Savage-Dickey density ratio (Heck, 2018; Verdinelli & Wasserman, 1995).

3 Results

Figure 1 shows the baseline probabilities p_i across the 16 studies and the corresponding empirically observed proportions of "yes"-responses. In all studies, these observed proportions were clearly below one, which means that a substantial number of participants responded honestly in each study. Moreover, except for one study (Moshagen, 2016), the empirical proportion of "yes"-responses exceeded the baseline probability p_i, implying that a substantial number of participants responded dishonestly. Figure 1 further demonstrates the importance of explicitly modeling the varying baseline probabilities p_i, as higher values on this design factor generally increased the probability of "yes"-responses. As a remedy, the extended regression approach developed above disentangles the varying baseline probabilities and the individual probabilities of being dishonest, which is not possible with existing statistical methods.

To obtain an unbiased estimate of the probability of dishonesty, we fitted the parameter d in the generalized version of the modified logistic regression with random intercepts. Across studies, the overall probability d was estimated to be 26.1%, with a 95% Bayesian credibility interval (BCI) of [19.2%; 34.4%]. Since our reanalysis focused on one-shot cheating paradigms, this implies that approximately one fourth of individuals were prepared to respond dishonestly. However, the uncertainty for the overall mean estimate was relatively large, which was due to the random-effect structure of the model, explicitly taking the heterogeneity of studies into account. Without random effects, the probability d was estimated to be 22.8% with a much narrower 95% BCI of [20.9%; 24.8%].

To facilitate the interpretation of effect sizes for the follow-

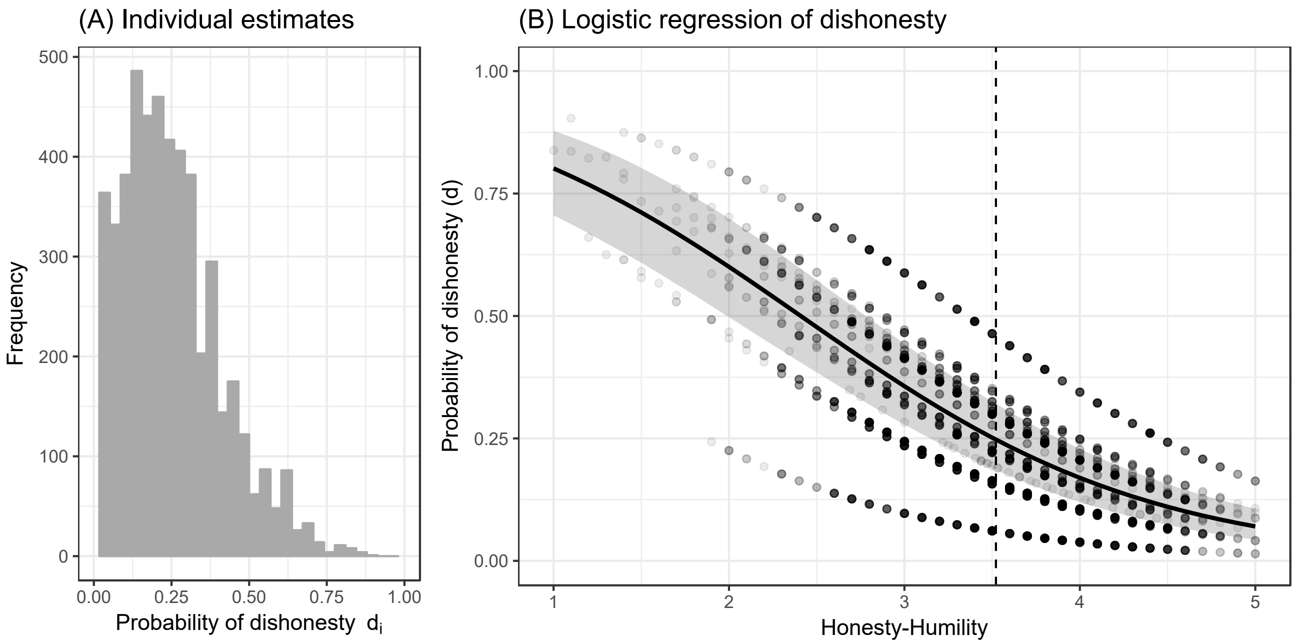

Figure 2: Panel A shows the individual estimates (posterior means) of the probability of dishonesty d_i. Panel B shows the same estimates (grey points) as a function of Honesty-Humility (HH), with the saturation indicating the number of participants. The dashed vertical line shows the overall mean of HH, and the solid curve the estimated logistic regression of d_i on the group level (with the 95% credibility interval in gray). Note that the individual parameter estimates are vertically scattered around this predicted curve due to the assumption of random-intercepts for the 16 datasets.

ing modified logistic regression models, we z-standardized predictors across studies (except when otherwise noted) and report odds ratios ($OR = \exp(\beta)$). The odds ratio for a z-standardized predictor quantifies how strongly the odds in favor of dishonesty change for individuals differing one SD on the predictor. To interpret effect sizes, we refer to odds ratios of OR = 1.25, 1.67, and 2.50 (and the inverse values 0.80, 0.60, and 0.40) for z-standardized predictors as small, medium, and large effects, respectively (these odds ratios approximately correspond to Cohen's d of 0.2, 0.5, and 0.8; Cohen, 1988).

3.1 Effect Size of Honesty-Humility

Our first goal was to provide a statistically precise effect size estimate of the HH-cheating link. HH resulted in a medium to large effect of OR = 0.53 with a 95% BCI of [0.47; 0.60], meaning that the odds of dishonesty are approximately halved for an increase of HH by one SD. The Bayes factor indicated evidence for a negative versus a null effect of HH on cheating ($B_{10} \approx 10^{26}$ one-sided). Given that Bayes factors larger than 100 are usually interpreted as extreme evidence for or against an effect (Jeffreys, 1961; Wagenmakers, Wetzels, Borsboom & Van Der Maas, 2011), this provides overwhelming evidence for a negative link between HH and dishonest behavior.

Figure 2A shows the distribution of the parameter estimates (posterior means) for the individual probabilities of dishonesty d_i (the corresponding 95% BCIs had a width of .129 on average with a maximum width of .357). As is apparent, the a-posteriori probability of dishonesty was below 50% for the majority of participants (i.e., 92.3%). In turn, Figure 2B shows these individual estimates for dishonesty as a function of individual levels in HH, demonstrating that, on average, the probability of dishonesty d decreased from 38.1% to 14.8% (53.6% to 8.5%) for individuals one (two) SD below versus above the mean of HH, respectively. As such, Figure 2B offers an alternative representation of the estimated effect size of OR = 0.53.

A possible concern might be that the effect-size estimate for the HH-cheating link is inflated due to consistent reporting. To test this alternative explanation, Hilbig and Zettler (2015) increased the temporal delay between the administration of the HEXACO-PI-R and the cheating paradigm to over two weeks, showing that the HH-cheating link remained significant. Since half of the studies in our reanalysis also implemented a temporal delay of at least two weeks (with a maximum of about half a year), we can apply a similar approach to test whether consistent responding can explain the moderate to strong effect of HH to dishonesty. For this purpose, we used a binary variable for studies without vs. with a delay of at least two weeks and fitted a modified logis-

https://doi.org/10.1017/S1930297500009232 Published online by Cambridge University Press

Table 2: Odds ratios for zero-order effects of the HEXACO factors, Honesty-Humility facets, and Big Five factors on dishonesty.

	Posterior of odds ratio			Bayes factor	
Covariate	Mean	2.5%	97.5%	B_{10}	B_{01}
HEXACO					
Honesty-Humility	0.53	0.47	0.60	$\approx \mathbf{10^{26}}$	$\approx 10^{-26}$
Emotionality	1.02	0.92	1.14	0.10	**10.30**
Extraversion	1.10	0.98	1.23	0.36	2.79
Agreeableness	0.91	0.82	1.01	0.39	2.54
Conscientiousness	0.86	0.77	0.96	2.79	0.36
Openness	0.93	0.83	1.03	0.22	**4.53**
Honesty-Humility facets					
Sincerity	0.66	0.59	0.74	$\approx \mathbf{10^{10}}$	$\approx 10^{-10}$
Fairness	0.63	0.56	0.70	$\approx \mathbf{10^{14}}$	$\approx 10^{-14}$
Greed Avoidance	0.67	0.60	0.75	$\approx \mathbf{10^{10}}$	$\approx 10^{-10}$
Modesty	0.64	0.58	0.71	$\approx \mathbf{10^{13}}$	$\approx 10^{-13}$
Big Five					
Neuroticism	1.18	0.98	1.42	0.73	1.36
Extraversion	1.11	0.91	1.36	0.27	**3.75**
Agreeableness	0.75	0.62	0.89	**40.49**	0.02
Conscientiousness	0.85	0.70	1.03	0.65	1.54
Openness	0.80	0.65	0.97	2.16	0.46

Note. Bayes factors larger than three indicate moderate evidence and are printed bold-faced. All Bayes factors are two-sided except for those related to HEXACO Honesty-Humility, the four Honesty-Humility facets, and Big Five Agreeableness (because for these traits, a directed negative effect was to be expected).

tic regression model that included two main effects for HH and delay as well as their interaction. The Bayes factor still showed clear evidence for the HH-cheating link ($B_{10} \approx 10^{26}$ with $OR = 0.53$; 95% BCI: [0.47; 0.60]), but ambiguous evidence for the main effect of temporal delay ($B_{10} = 1.0$; $OR = 0.71$ with 95% BCI: [0.44; 1.16]). Most importantly, the Bayes factor provided clear evidence against the one-sided prediction that the HH-cheating link decreases with temporal delay ($B_{01} = 18.0$), with an estimated odds ratio of $OR = 0.91$ for the interaction term (95% BCI: [0.77; 1.07]). Assuming that two weeks are sufficient to forget responses previously given in a personality questionnaire, these results imply that consistent responding cannot explain the moderate to strong effect of HH on dishonesty.

To further test which aspects of HH are predictive of dishonesty, we also ran (exploratory) analyses on the four facets of the HH factor (Lee & Ashton, 2004): Fairness (avoiding fraud and corruption), Sincerity (being genuine in interpersonal relations), Greed Avoidance (being uninterested in possessing luxury goods and signs of high social status), and Modesty (being modest and unassuming). Such a facet-level analysis is of interest because some of the HEXACO-PI-R items of the HH scale directly ask for dishonest behavior (e.g., "If I knew that I could never get caught, I would be willing to steal a million dollars"; Fairness facet) and may therefore primarily drive the HH-cheating link. Importantly, however, we relied on the HEXACO-60, which is not designed for facet-level analyses given that each facet is only measured by two to three items (Ashton & Lee, 2009). Hence, the following results should be interpreted with caution, as indicated by satisfactory but nevertheless rather low facet reliabilities (Sincerity: $\alpha = .60$; Fairness: $\alpha = .71$; Greed Avoidance: $\alpha = .61$; Modesty: $\alpha = .63$, in the current reanalysis). As summarized in Table 2, all four HH facets revealed small to medium-sized effects ranging between $OR = 0.63$ and $OR = 0.67$ and Bayes factors $B_{10} > 10^{10}$, supporting a negative association of each facet with unethical decision making. Moreover, each facet had incremental validity in a regression model that included all four facet-level scales as predictors. The estimated effect sizes were $OR = 0.84$ for Sincerity (95% BCI: [0.74, 0.95]; $B_{10} = 16.2$), $OR = 0.75$ for Fairness (95% BCI: [0.67, 0.85]; $B_{10} \approx 7{,}800$), $OR = 0.84$ for Greed Avoidance (95% BCI: [0.74, 0.95]; $B_{10} = 15.1$) and $OR = 0.75$ for Modesty (95% BCI: [0.66, 0.83]; $B_{10} \approx 10^4$). These findings – at least tentatively – support that each of the four HH facets explains unique variance in unethical decision making, thus implying that the effect of HH on dishonesty cannot be merely attributed to a few HEXACO-PI-R items being directly related to cheating.

3.2 Effect Size and Incremental Validity of Other Basic Personality Traits

As for HH, we next fitted separate modified logistic regression models to estimate the zero-order effects of other basic personality traits from the HEXACO model and the Big Five. These analyses used only a subset of the data for which information on the corresponding traits was available (i.e., $N = 4{,}965$ responses for the HEXACO and $N = 1{,}650$ for the Big Five). As is apparent in Table 2, none of the HEXACO traits apart from HH had a noteworthy association with dishonesty. Only the Bayes factor for Conscientiousness ($B_{10} = 2.8$) indicated anecdotal evidence for an (undirected) effect versus a null effect. Moreover, the Bayes factor indicated moderate evidence for the absence of an effect for the HEXACO dimensions Emotionality ($B_{01} = 10.3$) and Openness to Experience ($B_{01} = 4.5$). In contrast, the Bayes factors for Extraversion ($B_{01} = 2.8$) and Agreeableness ($B_{01} = 2.5$) indicated only ambiguous evidence against an undirected effect.

Table 3: Incremental validity of the HEXACO and Big Five factors, respectively, above and beyond Honesty-Humility.

Covariate	Posterior of odds ratio Mean	2.5%	97.5%	Bayes factor B_{10}	B_{01}
HEXACO					
Honesty-Humility	0.53	0.46	0.60	$\approx \mathbf{10^{24}}$	$\approx 10^{-24}$
Emotionality	1.04	0.93	1.17	0.20	**4.92**
Extraversion	1.14	1.01	1.30	1.46	0.68
Agreeableness	1.09	0.97	1.23	0.43	2.35
Conscientiousness	0.94	0.84	1.05	0.29	**3.50**
Openness	0.94	0.83	1.05	0.31	**3.25**
Big Five					
Honesty-Humility	0.54	0.42	0.69	$\approx \mathbf{10^{5}}$	$\approx 10^{-5}$
Neuroticism	1.14	0.91	1.43	0.50	2.01
Extraversion	1.08	0.86	1.37	0.33	**3.06**
Agreeableness	0.97	0.78	1.21	0.26	**3.92**
Conscientiousness	0.94	0.76	1.14	0.31	**3.27**
Openness	0.79	0.64	0.97	**3.48**	0.29

Note. Bayes factors larger than three indicate moderate evidence and are printed bold-faced. All Bayes factors are two-sided except for those related to Honesty-Humility and Big Five Agreeableness (because for these traits, a directed negative effect was to be expected). The Bayes factor testing a model with Honesty-Humility only compared to the full model was B_{01} = 412.7 for the HEXACO and B_{01} = 56.7 for the Big Five.

Among the Big Five dimensions, in turn, we found strong evidence (B_{10} = 40.5, one-sided) for a negative zero-order effect (versus a null effect) of Agreeableness on dishonesty. The corresponding odds ratio indicated a small to medium-sized effect, *OR* = 0.75, 95% BCI [0.62; 0.89], showing that the odds of dishonesty reduced by approximately a fourth for an increase of Big Five Agreeableness by one SD. However, the link of Big Five Agreeableness to dishonesty was considerably smaller than the corresponding link of HH as indicated by non-overlapping credibility intervals for the zero-order effects. For the remaining Big Five dimensions, evidence was ambiguous, except for Extraversion, for which the Bayes factor indicated moderate evidence for the absence of an effect (B_{01} = 3.8). The Bayesian credibility interval indicated a small trend that higher levels of Big Five Openness were associated with lower dishonesty (*OR* = 0.80, 95% BCI [0.65; 0.97]), but the corresponding Bayes factor provided only anecdotal evidence for this effect (B_{10} = 2.2).

Next, we fitted a regression model including all six HEXACO traits as predictors to test whether other HEXACO traits have any incremental validity above HH for the prediction of dishonesty. As is apparent in Table 3, HH was still the single most valid predictor of dishonesty ($B_{10} \approx 10^{24}$), revealing an identical effect size as compared to its zero-order effect (*OR* = 0.53, 95% BCI [0.46; 0.60]). Unlike the zero-order effects, it is noteworthy that the Bayes factor now provided moderate evidence for null effects of Conscientiousness (B_{01} = 3.5) and Openness (B_{01} = 3.3). For the remaining HEXACO dimensions, the Bayes factor again indicated evidence for a null effect of Emotionality (B_{01} = 4.9), but only ambiguous evidence for whether Extraversion (B_{10} = 1.5) or Agreeableness (B_{10} = 0.4) contributed unique variance above and beyond the other traits. To test the joint incremental validity of the five remaining HEXACO traits beyond HH, we compared the full model including all six HEXACO traits as predictors against a model including HH only. The Bayes factor clearly favored the HH-only model (B_{01} = 412.7), thus indicating very strong evidence against the incremental validity of the other five HEXACO traits beyond HH.

Table 3 also summarizes the estimates for a model including HH along with the Big Five dimensions. As before, we found clear evidence that HH had a negative effect ($B_{10} \approx 10^5$), with a highly similar effect size as the zero-order effect (*OR* = 0.54, 95% BCI: [0.42; 0.69]). Moreover, the Bayes factor indicated moderate evidence for the absence of an effect of Big Five Agreeableness (B_{01} = 3.9). This is especially noteworthy given the substantial zero-order effect of Big Five Agreeableness (Table 2). Importantly, this result cannot be attributed to the other four Big Five traits as it replicated when fitting a model including only HH and Big Five Agreeableness as predictors. In this model, too, the Bayes factor provided moderate evidence against the incremental validity of Big Five Agreeableness (B_{01} = 4.0), but still very strong evidence for the negative effect of HH ($B_{10} \approx 10^6$). In turn, for the remaining Big Five traits in the Big Five plus HH model, the Bayes factors indicated slight evidence for the incremental validity of Openness (B_{10} = 3.5) – implying a weak negative effect (*OR* = 0.79, 95% BCI [0.64; 0.97]) –, slight evidence against the incremental validity of Extraversion (B_{01} = 3.1) and Conscientiousness (B_{01} = 3.3), and ambiguous evidence with respect to the incremental validity of Neuroticism (B_{01} = 2.0). Importantly, overall, the Big Five traits showed no incremental validity above a model including HH only (B_{01} = 56.7).

3.3 Honesty-Humility and Demographic Characteristics

To test whether the effect of HH diminishes or disappears when controlling for demographic characteristics, we fitted a model with HH and participants' sex and age as covariates (effect-coded and centered, respectively). For this analysis, we excluded 10 participants who did not provide their sex or age, which led to a total sample size of *N* = 4,992 for this analysis.

According to the Bayes factor, participants' sex had no effect on dishonest behavior (B_{01} = 8.1). However, age (measured in years) was a valid predictor of dishonesty (B_{10} = 29.1) with an odds ratio of *OR* = 0.985 and a 95% BCI of [0.976; 0.994].[5] Substantively, this implies that the odds of dishonest behavior were 0.857 and 0.735 times smaller for individuals 10 and 20 years older than any reference group, respectively. To test the alternative explanation that this age effect was simply due to different levels of dishonesty between student and community samples (the mean age for student samples was lower than for community samples; Table 1), we also fitted a model that included this variable as an additional, binary predictor. However, the distinction between student and community samples did not provide any incremental validity (B_{01} = 2.2) and the Bayes factor in favor of an effect of age remained robust (B_{10} = 31.3). More importantly, the Bayes factor indicated overwhelming evidence for the incremental validity of HH over and above the demographic characteristics ($B_{10} \approx 10^{22}$), with a highly similar odds ratio as observed for the zero-order effect (*OR* = 0.56, 95% BCI [0.49; 0.62]). Thus, the widely replicated link between HH and dishonest behavior cannot be attributed to the covariation between HH and sex or age.

3.4 Interaction of HH with Other Traits

To test whether HEXACO Conscientiousness or Emotionality moderate the HH-cheating link, we fitted two models, each including the main effect of HH, the main effect of one of the two other traits of interest, and their multiplicative interaction term. Given that HEXACO theory provides directed (disordinal) hypotheses regarding potential interaction effects of HH with Conscientiousness and Emotionality, respectively (see above), these were tested using one-sided Bayes factors. For the HH*Conscientiousness model, the Bayes factor showed moderate evidence for the absence of both a main effect of Conscientiousness and a disordinal interaction with HH (B_{01} = 7.5 two-sided and B_{01} = 14.0 one-sided, respectively). By contrast, the negative main effect of HH remained robust with a similar, medium to large effect size as before (*OR* = 0.54, 95% BCI [0.47; 0.60], $B_{10} \approx 10^{25}$ one-sided). For the HH*Emotionality model, in turn, the Bayes factor likewise indicated evidence for a null effect of Emotionality (B_{01} = 5.9 two-sided) and the absence of a negative effect of the interaction of Emotionality and HH (B_{01} = 23.0 one-sided). The corresponding estimated odds ratio of the interaction was *OR* = 1.10 with a 95% BCI of [0.99; 1.22], thus even showing a negligible trend in the opposite direction as predicted.

[5]Since age was centered but not *z*-standardized, this odds ratio must not be interpreted using the conventions for small, medium, and large effects from above.

3.5 Boundary Conditions of the HH-Cheating Link

Finally, to investigate the boundary conditions of the HH-cheating link, we tested whether the baseline probability *p* in the cheating paradigm moderates the effect of HH on dishonesty. The corresponding modified logistic regression model included main effects for HH and the baseline probability *p* as well as a multiplicative interaction term. We excluded one study in which participants were unaware of the actual baseline probability *p* because the die was biased (Perugini & Leone, 2009), resulting in *N* = 4,965 responses for this analysis. Replicating our prior findings, HH again had a large negative effect on dishonesty ($B_{10} \approx 10^{26}$) with a medium to large odds ratio of *OR* = 0.53 and a 95% BCI of [0.47; 0.60]. However, evidence for the main effect of the baseline probability *p* (which was predicted to be positive) remained ambiguous (B_{01} = 1.4 one-sided), whereas the Bayes factor for the interaction of *p* with HH showed clear evidence against the predicted negative effect (B_{01} = 14.5 one-sided).

4 Discussion

To explain the substantial individual differences in dishonest behavior, research has focused on linking unethical decisions in cheating paradigms to basic personality traits in general and the Honesty-Humility (HH) dimension of the HEXACO personality model in particular (e.g., Hilbig et al., 2015; Kleinlogel et al., 2018; Perugini & Leone, 2009; Thielmann et al., 2017). Once viewed through the lens of cumulative science and the importance of estimating effect sizes, several core questions have remained largely unanswered, partly due to small samples and, by implication, insufficient statistical power. As a remedy, we reanalyzed the raw data of 16 studies with a total sample size of *N* = 5,002 to (1) provide an unbiased and statistically more precise estimate of the effect linking HH and dishonesty, (2) clarify whether other basic traits account for dishonesty in addition to and/or above HH, (3) test whether the HH-dishonesty link is partially or completely attributable to demographic characteristics (sex or age), (4) investigate potential interaction effects between HH and other basic traits, and (5) examine whether the baseline probability *p* of the cheating paradigm moderates the HH-cheating link.

Overall, the reanalysis provided overwhelming evidence that HH is the single most valid predictor of dishonest behavior amongst basic personality traits. The medium to large effect size of this link is best quantified as an odds ratio: The odds in favor of dishonesty in a cheating paradigm are (approximately) doubled for individuals who are one standard deviation lower in HH (*OR* = 0.53). In other words, the probability of dishonesty, *d*, decreased from 38.1% to 14.8% for individuals one SD below versus above the mean of HH (Figure 2). Besides HH, only Big Five Agreeableness showed a small (negative) zero-order effect on dishonesty, which however disappeared in a multiple regression model including HH as an additional predictor. Indeed, none of the remaining HEXACO or the Big Five traits showed any incremental validity above HH. This once more supports the theoretical conceptualization of the HH dimension and its incremental validity beyond the Big Five traits (Ashton & Lee, 2005, 2008b, 2008a).

Moreover, our analysis revealed that the HH-cheating link is not merely attributable to correlations between HH and sex or age, respectively (Ashton & Lee, 2016; Lee & Ashton, in press; Moshagen, Hilbig & Zettler, 2014). Although participants' age – but not their sex – explained unique variance in dishonesty above HH, the effect of HH remained stable and similar in size when considering sex and age as additional predictors. Note that our results showed no effect of sex, but implied that older individuals were less likely to cheat than younger individuals. This is in contrast to recent meta-analytic evidence (Abeler et al., 2016) showing no effect of age but of sex on dishonesty – with women cheating less than men (however, some studies showed a negative effect of age; e.g., Bucciol et al., 2013).

The finding that HH in and of itself accounts for individual variation in dishonesty may not be surprising given that some of the HH items of the HEXACO-PI-R directly ask for criminal and unethical tendencies. Thus, from a theoretical perspective, the explanandum (dishonesty) is very close to the explanans (items assessing honest and prosocial behavior). However, despite this conceptual proximity of the theoretical definition and operationalization of HH to cheating behavior, it is important to test empirically whether HH (a trait measured by a self-report questionnaire) is predictive of actual, incentivized unethical decisions (a behavioral variable observed in the cheating paradigm). Moreover, (exploratory) facet-level analyses showed that each of the four HH facets is negatively linked to dishonesty and has incremental validity beyond the other facets. This suggests that the HH-cheating link cannot merely be attributed to a few specific items. Also, the effect is unlikely to be driven by consistent responding, since time gaps of two weeks and more between administration of the HEXACO-PI-R and the cheating task did not moderate the HH-cheating link.

We also found evidence against theoretically derived interactions of HH with other personality traits (i.e., Conscientiousness and Emotionality; Lee & Ashton, 2013). First, according to HEXACO theory, individuals low in both HH and Conscientiousness should care less about (ethical) rules and be more impulsive, which would arguably increase the odds of dishonesty in cheating paradigms (see also Shalvi et al., 2012). However, we found clear evidence against such an interaction, suggesting that individuals simply follow their inclination to behave dishonestly in cheating paradigms while neglecting their inclination for rule compliance versus impulsivity. Second, individuals low in both HH and Emotionality should be particularly inclined towards dishonesty since they will fear the consequences of cheating less than their counterparts high in Emotionality. In contrast to this prediction, we found evidence against both a main effect of Emotionality and an interaction with HH. Given that Emotionality covers individual differences in trait anxiety, this null effect stands in contrast to recent arguments claiming that anxiety will foster unethical behavior (Lu, Lee, Gino & Galinsky, 2018). Nonetheless, the finding can be plausibly attributed to the fact that, as intended, cheating paradigms reduce (or even fully remove) the fear of being exposed as a cheater and facing corresponding sanctions by others (Abeler et al., in press).

Finally, our analysis also provided evidence against an interaction between HH and a crucial design factor of the cheating paradigm, the baseline probability p of obtaining the favorable outcome (thus replicating a previous finding by Hilbig & Zettler, 2015). That is, although a lower baseline probability increases the suspiciousness of "yes"-responses and thus renders them more self-incriminating, individuals low in HH were as willing – and those high in HH as unwilling – to lie when the baseline probability was $p = .161$ as compared to $p = .375$. In this regard, our results also showed evidence against a main effect of the baseline probability p, meaning that the probability of legitimately responding "yes" did not affect the probability of dishonesty. On the one hand, these results might be attributable to the relatively small variation in p across studies included in the current reanalysis (Figure 1) – as also implied by the findings of Abeler and colleagues (2016) that the baseline probability (varying between $p = .10$ versus $p = .60$) indeed positively affected the probability of dishonesty. On the other hand, they suggest that individuals felt their privacy to be equally well protected – as intended by the design of the cheating paradigm.

4.1 Methodological Developments

To allow for the reanalysis of raw data across studies, we developed a generalized version of the modified logistic regression for the cheating paradigm (Moshagen & Hilbig, 2017). The new extension permits baseline probabilities p varying across observations and accounts for nested data structures (e.g., participants nested within studies or repeated responses nested within participants). Moreover, our statistical conclusions relied on Bayes factors, which quantify the evidence in favor of a null hypothesis and against an alternative. Based on this method, the reanalysis provided clear evidence *against* the incremental validity of any other traits beyond HH, and *for* the hypothesis that the link between HH and dishonesty is invariant in the context of other basic traits and demographic variables (i.e., sex and age), and neither moderated by other basic traits nor by the baseline probability p of the cheating paradigm. In general, note that the Bayes factor takes the sample size into account, which is why our results cannot be merely attributed to limited statistical power to detect small effects (Wagenmakers, 2007; see also sensitivity simulation in the Appendix).[6]

The generalization of the modified logistic regression allows testing novel types of hypotheses on the link between dishonesty and personality traits, situational factors, crucial design factors, and their interactions. Another important application of the generalized framework refers to within-subject designs. For instance, in the study by Klein et al. (2017), participants played the dice-roll task six times with different payoff structures serving as a within-subject factor. To test their theoretical predictions while simultaneously accounting for the within-subject structure, Klein et al. (2017) used a hierarchical version of a standard logistic regression to model the observable "yes"-responses. However, as explained above, this approach will result in underestimation of effect sizes. Therefore, the statistical analysis needs to combine a modified logistic regression for the cheating paradigm with random intercepts for the within-subject design to account for the nested data structure while nonetheless allowing provision of unbiased effect size estimates. The newly developed generalized regression model fulfills both these requirements. As an alternative to the Bayesian implementation in the present paper, note that the hierarchical version of the modified regression can also be fitted with the RRmixed function of the R package RRreg (Heck & Moshagen, 2018) which provides maximum-likelihood estimates if the baseline probability p is constant across all observations.

4.2 Limitations

Although the current reanalysis provides vital insights on individual differences in ethical decision making, some limitations should be acknowledged. First, our reanalysis focused on studies that used a specific type of paradigm for measuring dishonesty, namely, the dice-roll or coin-toss task. Although HH has also been found to predict cheating in an anagram task (Hershfield, Cohen & Thompson, 2012; Study 4) or in a self-scoring knowledge task (Hilbig & Zettler, 2015; Study 1), it remains an open question whether the HH-cheating link is as robust in these other paradigms as in the ones focused on herein. Second, given that our reanalysis concentrated on basic personality traits from the HEXACO model and the Big Five, it remains unclear whether HH has incremental validity over more narrow traits such as the Dark Triad traits (Paulhus & Williams, 2002) or the dark factor of personality, D (Moshagen et al., in press). Even though HH is closely linked to the Dark Triad (e.g., Lee & Ashton, 2005, 2014) and also has a latent correlation of $\hat{\rho}$ = .80 with D, the factor D has already been shown to be theoretically and empirically distinguishable from HH (Moshagen et al., in press). Therefore, it remains an open question for future research whether HH also explains additional variance over and above these specific dark traits.

Another limitation is that only some of the studies included in the reanalysis used the NEO-FFI to measure the Big Five. Due to the smaller total sample size, conclusions for the Big Five dimensions were more ambiguous than those for the HEXACO traits. For the available sample size of N = 1,650 observations for the Big Five, the sensitivity simulation in the Appendix shows that the distribution of simulated Bayes factors has a noteworthy probability of being ambiguous (i.e., B_{10} between 1/3 and 3) if the true effect is small. Future meta-analytic research would therefore require an even larger number of observations than our reanalysis to unambiguously distinguish between null effects and small effects of the Big Five traits. According to the simulation, doubling the sample size to N = 3,200, for instance, would considerably raise the probability of finding moderate evidence even for a small effect (OR = 1.25) to 75%.

Also, the new statistical framework is currently restricted to the binary cheating paradigm, which in turn limits the range of substantive hypotheses that can be addressed. If each participant only provides a single response, it is not possible to distinguish between "moderate" and "extreme" cheaters or to test hypotheses concerning the temporal dynamics of cheating behavior in repeated decisions at an individual level (Hilbig & Thielmann, 2017). However, the modified logistic regression approach can be adapted for a paradigm in which participants report the total number of "yes"-responses across multiple dice rolls instead of a dichotomous "yes"/"no"-response (Kleinlogel et al., 2018; Schindler & Pfattheicher, 2017). In this case, the dependent variable "number of yes-responses" ranges between y_i = 0 and the maximum number of "yes"-responses possible, $y_i = n_i$, and thus responses can be modeled by a binomial distribution with success probability identical to Equation 1. Note, however, that each individual still provides only one value on the HH scale and only one (non-binary) value in the cheating paradigm. Since the binomial distribution assumes that responses correspond to n_i independent observations, it is necessary to reweigh the responses or to fit a hierarchical model that explicitly models the frequencies as nested within individuals and uses personality traits as level-2 predictors (Heck et al., 2018).

4.3 Conclusions

To conclude, we found strong evidence that HH is the most consistent predictor of dishonest behavior among basic personality traits. That is, the HH-cheating link was extremely

[6] This logic of providing evidence for the null hypothesis is an advantage of the Bayesian framework, but it is not a unique feature. In the classical Neyman-Pearson framework, the null hypothesis is tested against a precisely defined alternative hypothesis, which in turn allows researchers to decide *for* or *against* the null hypothesis (Moshagen & Erdfelder, 2016).

robust across regression models that showed neither main effects nor interactions of any other basic trait dimension from the HEXACO or Big Five model. This invariance of the HH-dishonesty link is theoretically important because it supports a simple and parsimonious explanation of individual differences in dishonest behavior. Put provocatively, our results are especially relevant vis-à-vis Meehl's (1990, p. 204) crud factor, which states that "everything correlates to some extent with everything else" in the social sciences, including personality psychology. From this perspective, providing evidence for a robust and invariant link between a basic trait dimension and unethical decision making contributes more to a cumulative science than establishing complex patterns of correlations and regression weights that merely reflect the crud factor.

References

Abeler, J., Nosenzo, D., & Raymond, C. (in press). Preferences for truth-telling. *Econometrica*. Retrieved from http://EconPapers.repec.org/RePEc:not:notcdx:2016-13.

Ashton, M. C., & Lee, K. (2005). Honesty-Humility, the Big Five, and the Five-Factor Model. *Journal of Personality*, *73*(5), 1321–1354. http://dx.doi.org/10.1111/j.1467-6494.2005.00351.x.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*(2), 150–166. http://dx.doi.org/10.1177/1088868306294907.

Ashton, M. C., & Lee, K. (2008a). The HEXACO model of personality structure and the importance of the H factor. *Social and Personality Psychology Compass*, *2*(5), 1952–1962. http://dx.doi.org/10.1111/j.1751-9004.2008.00134.x.

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

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

Ashton, M. C., & Lee, K. (2016). Age trends in HEXACO-PI-R self-reports. *Journal of Research in Personality*, *64*, 102–111. http://dx.doi.org/10.1016/j.jrp.2016.08.008.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396–403. http://dx.doi.org/10.1111/j.1745-6916.2007.00051.x.

Bucciol, A., Landini, F., & Piovesan, M. (2013). Unethical behavior in the field: Demographic characteristics and beliefs of the cheater. *Journal of Economic Behavior & Organization*, *93*, 248–257. http://dx.doi.org/10.1016/j.jebo.2013.03.018.

Bucciol, A., & Piovesan, M. (2011). Luck or cheating? A field experiment on honesty with children. *Journal of Economic Psychology*, *32*(1), 73–78. http://dx.doi.org/10.1016/j.joep.2010.12.001.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. http://dx.doi.org/10.18637/jss.v076.i01.

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

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

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. http://dx.doi.org/10.1177/0956797613504966.

Dana, J., Weber, R., & Kuang, J. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80. http://dx.doi.org/10.1007/s00199-006-0153-z.

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525–547. http://dx.doi.org/10.1111/jeea.12014.

Gronau, Q. F., Wagenmakers, E.-J., Heck, D. W., & Matzke, D. (2017). A simple method for comparing complex models: Bayesian model comparison for hierarchical multinomial processing tree models using warp-III bridge sampling. Manuscript under revision. Retrieved from https://psyarxiv.com/yxhfm/.

Heck, D. W. (2018). *A caveat on the Savage-Dickey density ratio: The case of computing Bayes factors for regression parameters*. Manuscript under revision.

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*(1), 264–284. http://dx.doi.org/10.3758/s13428-017-0869-7.

Heck, D. W., & Moshagen, M. (2018). RRreg: An R package for correlation and regression analyses of randomized response data. *Journal of Statistical Software*, *85(2)*, 1–29. http://dx.doi.org/10.18637/jss.v085.i02.

Hershfield, H. E., Cohen, T. R., & Thompson, L. (2012). Short horizons and tempting situations: Lack of continuity to our future selves leads to unethical decision making and behavior. *Organizational Behavior and Human Decision*

Processes, *117*(2), 298–310. http://dx.doi.org/10.1016/j.obhdp.2011.11.002.

Hilbig, B. E., & Hessler, C. M. (2013). What lies beneath: How the distance between truth and lie drives dishonesty. *Journal of Experimental Social Psychology*, *49*(2), 263–266. http://dx.doi.org/10.1016/j.jesp.2012.11.010.

Hilbig, B. E., Moshagen, M., & Zettler, I. (2015). Truth will out: Linking personality, morality, and honesty through indirect questioning. *Social Psychological and Personality Science*, *6*(2), 140–147. http://dx.doi.org/10.1177/1948550614553640.

Hilbig, B. E., Moshagen, M., & Zettler, I. (2016). Prediction consistency: A test of the equivalence assumption across different indicators of the same construct. *European Journal of Personality*, *30*(6), 637–647. http://dx.doi.org/10.1002/per.2085.

Hilbig, B. E., & Thielmann, I. (2017). Does everyone have a price? On the role of payoff magnitude for ethical decision making. *Cognition*, *163*, 15–25. http://dx.doi.org/10.1016/j.cognition.2017.02.011.

Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. *Journal of Research in Personality*, *57*, 72–88. http://dx.doi.org/10.1016/j.jrp.2015.04.003.

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, *21*(5), 1157–1164. http://dx.doi.org/10.3758/s13423-013-0572-3.

Horn, J., Nelson, C. E., & Brannick, M. T. (2004). Integrity, conscientiousness, and honesty. *Psychological Reports*, *95*(1), 27–38. http://dx.doi.org/10.2466/pr0.95.1.27-38.

Jeffreys, H. (1961). *Theory of probability*. New York: Oxford University Press.

Jones, D. N., & Paulhus, D. L. (2017). Duplicity among the dark triad: Three faces of deceit. *Journal of Personality and Social Psychology*, *113*(2), 329–342. http://dx.doi.org/10.1037/pspp0000139.

Kajackaite, A., & Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior*, *102*, 433–444. http://dx.doi.org/10.1016/j.geb.2017.01.015.

Klein, S. A., Thielmann, I., Hilbig, B. E., & Zettler, I. (2017). Between me and we: The importance of self-profit versus social justifiability for ethical decision making. *Judgment and Decision Making*, *12*(6), 563–571.

Klein, S. A., Thielmann, I., Hilbig, B. E., & Zettler, I. (2018). *Unpublished dataset*.

Kleinlogel, E. P., Dietz, J., & Antonakis, J. (2018). Lucky, competent, or just a cheat? Interactive effects of honesty-humility and moral cues on cheating behavior. *Personality and Social Psychology Bulletin*, *44*(2), 158–172. http://dx.doi.org/10.1177/0146167217733071.

Lee, K., & Ashton, M. C. (in press). Psychometric properties of the HEXACO-100. *Assessment*, 1073191116659134. http://dx.doi.org/10.1177/1073191116659134.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. http://dx.doi.org/10.1207/s15327906mbr3902_8.

Lee, K., & Ashton, M. C. (2005). Psychopathy, Machiavellianism, and Narcissism in the Five-Factor Model and the HEXACO model of personality structure. *Personality and Individual Differences*, *38*(7), 1571–1582. http://dx.doi.org/10.1016/j.paid.2004.09.016.

Lee, K., & Ashton, M. C. (2013). *The H factor of personality: Why some people are manipulative, self-entitled, materialistic, and exploitive—and why it matters for everyone*. Waterloo, Ontario: Wilfrid Laurier.

Lee, K., & Ashton, M. C. (2014). The Dark Triad, the Big Five, and the HEXACO model. *Personality and Individual Differences*, *67*, 2–5. http://dx.doi.org/10.1016/j.paid.2014.01.048.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g-priors for Bayesian variable selection. *Journal of the American Statistical Association*, *103*(481), 410–423. http://dx.doi.org/10.1198/016214507000001337.

Lu, J. G., Lee, J. J., Gino, F., & Galinsky, A. D. (2018). Polluted morality: Air pollution predicts criminal activity and unethical behavior. *Psychological Science*, *29*(3), 340–355. http://dx.doi.org/10.1177/0956797617735807.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633–644. http://dx.doi.org/10.1509/jmkr.45.6.633.

Mazar, N., & Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy & Marketing*, *25*(1), 117–126.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*(1), 195–244. http://dx.doi.org/10.2466/pr0.1990.66.1.195.

Moshagen, M. (2016). *Unpublished dataset*.

Moshagen, M., & Erdfelder, E. (2016). A new strategy for testing structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 54–60. http://dx.doi.org/10.1080/10705511.2014.950896.

Moshagen, M., & Hilbig, B. E. (2017). The statistical analysis of cheating paradigms. *Behavior Research Methods*, *49*(2), 724–732. http://dx.doi.org/10.3758/s13428-016-0729-x.

Moshagen, M., Hilbig, B. E., Erdfelder, E., & Moritz, A. (2014). An experimental validation method for questioning techniques that assess sensitive issues. *Experimental Psychology*, *61*(1), 48–54. http://dx.doi.org/10.1027/1618-3169/a000226.

Moshagen, M., Hilbig, B. E., & Zettler, I. (in press). The dark core of personality. *Psychological Review*. http://dx.doi.org/10.1037/rev0000111.

Moshagen, M., Hilbig, B. E., & Zettler, I. (2014). Faktorenstruktur, psychometrische Eigenschaften und Messinvarianz der deutschsprachigen Version des 60-Item HEXACO Persönlichkeitsinventars. *Diagnostica*, *60*, 86–97. http://dx.doi.org/10.1026/0012-1924/a000112.

Müller, S., & Moshagen, M. (2018). True virtue, self-presentation, or neither? A behavioral test of impression management and overclaiming. *Submitted for Publication*.

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556–563. http://dx.doi.org/10.1016/S0092-6566(02)00505-6.

Perugini, M., & Leone, L. (2009). Implicit self-concept and moral action. *Journal of Research in Personality*, *43*(5), 747–754. http://dx.doi.org/10.1016/j.jrp.2009.03.015.

Pfattheicher, S., Schindler, S., & Nockur, L. (in press). On the impact of Honesty-Humility and a cue of being watched on cheating behavior. *Journal of Economic Psychology*. http://dx.doi.org/10.1016/j.joep.2018.06.004.

Rouder, J. N., & Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research*, *47*, 877–903. http://dx.doi.org/10.1080/00273171.2012.734737.

Schindler, S., & Pfattheicher, S. (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology*, *69*, 172–177. http://dx.doi.org/10.1016/j.jesp.2016.09.009.

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science*, *23*(10), 1264–1270. http://dx.doi.org/10.1177/0956797612443835.

Shalvi, S., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). Ethical manoeuvring: Why people avoid both major and minor lies. *British Journal of Management*, *22*, S16–S27. http://dx.doi.org/10.1111/j.1467-8551.2010.00709.x.

Thielmann, I., & Hilbig, B. E. (in press). No gain without pain: The psychological costs of dishonesty. *Journal of Economic Psychology*. http://dx.doi.org/10.1016/j.joep.2018.06.001.

Thielmann, I., Hilbig, B. E., Zettler, I., & Moshagen, M. (2017). On measuring the sixth basic personality dimension: A comparison between HEXACO honesty-humility and big six honesty-propriety. *Assessment*, *24*(8), 1024–1036. http://dx.doi.org/10.1177/1073191116638411.

Thielmann, I., Klein, S. A., & Hilbig, B. E. (2018). Who cheats for whom? *Manuscript in Preparation*.

Ulrich, R., Schröter, H., Striegel, H., & Simon, P. (2012). Asking sensitive questions: A statistical power analysis of randomized response models. *Psychological Methods*, *17*(4), 623–641. http://dx.doi.org/10.1037/a0029314.

Utikal, V., & Fischbacher, U. (2013). Disadvantageous lies in individual decisions. *Journal of Economic Behavior & Organization*, *85*, 108–111. http://dx.doi.org/10.1016/j.jebo.2012.11.011.

Verdinelli, I., & Wasserman, L. (1995). Computing Bayes Factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association*, *90*(430), 614–618. http://dx.doi.org/10.1080/01621459.1995.10476554.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. http://dx.doi.org/10.3758/BF03194105.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & Van Der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*(3), 426-432. http://dx.doi.org/10.1037/a0022790.

Williams, K. M., Nathanson, C., & Paulhus, D. L. (2010). Identifying and profiling scholastic cheaters: Their personality, cognitive ability, and motivation. *Journal of Experimental Psychology: Applied*, *16*(3), 293–307. http://dx.doi.org/10.1037/a0020773.

Zettler, I., Hilbig, B. E., Moshagen, M., & de Vries, R. E. (2015). Dishonest responding or true virtue? A behavioral test of impression management. *Personality and Individual Differences*, *81*, 107–111. http://dx.doi.org/10.1016/j.paid.2014.10.007.

Appendix: Sensitivity Simulation for the Bayes Factor

Previous studies investigating the link between personality traits and dishonesty in cheating paradigms used relatively small samples, resulting in a low sensitivity to distinguish between small and null effects. To illustrate this in the Bayesian framework, we simulated the distribution of Bayes factors for various sample and effect sizes. We varied the sample size between N = 100, N = 200, N = 400, etc. up to N = 6,400. Moreover, we generated values for a standard-normally distributed variable with an odds ratio of 1.00, 1.25, and 1.67 of the effect on dishonesty, reflecting a zero, small, and medium effect size, respectively. We used a fixed baseline probability of p = .25, the probability used in most of the studies included in our reanalysis. Note that smaller (larger) baseline probabilities p result in higher (lower) statistical power (Ulrich, Schröter, Striegel, & Simon, 2012). Moreover, we used a fixed-effects model for data generation and analysis, assuming that all responses are independent and identically distributed.

Figure A1 shows the distribution of the Bayes factor for different numbers of observations and effect sizes. In the

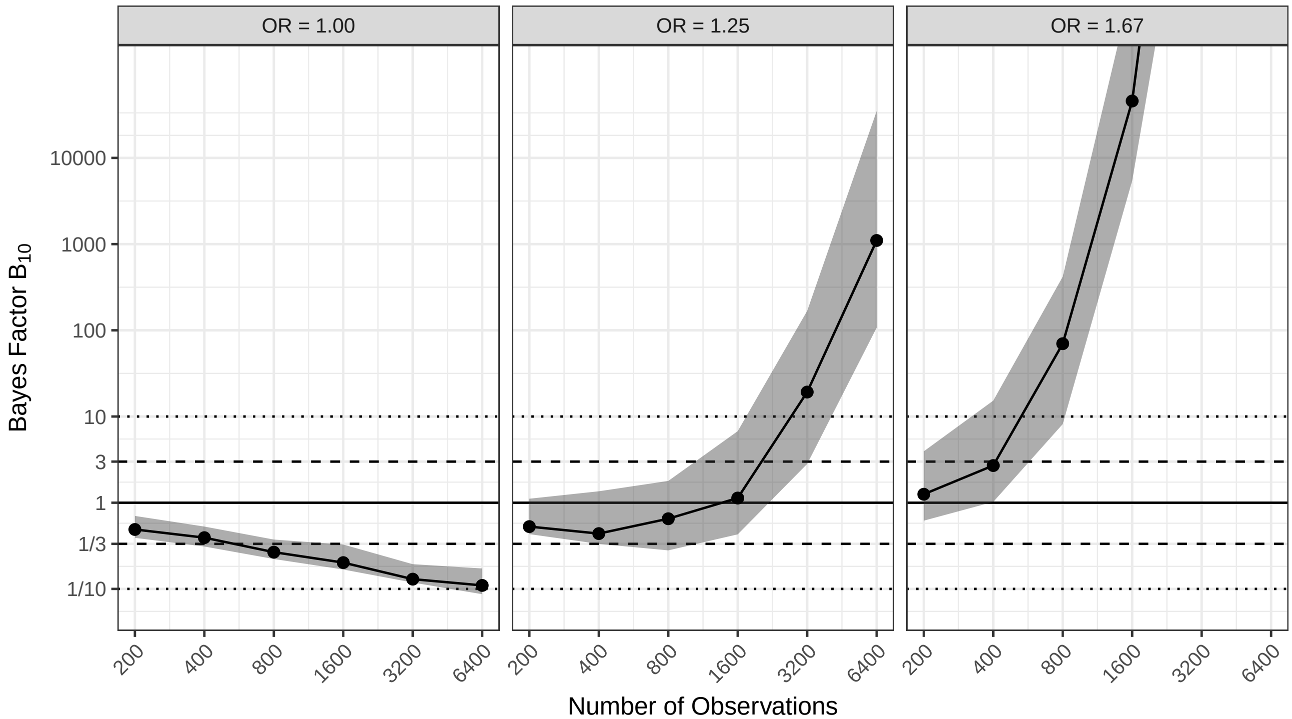

Figure A1. Distribution of 100 simulated Bayes factors as a function of sample size for a zero (left panel), small (middle panel), and medium effect size (right panel) of a z-standardized predictor. The black points connected by solid lines show the median Bayes factor, and the gray ribbon the 25% and 75% quantiles. The dashed and dotted horizontal lines show the conventional boundaries of 3 and 10 for moderate and strong evidence, respectively. Note that log-scales are used for the Bayes factor and the sample size to facilitate readability.

middle and right panel, the predictor has an effect on dishonesty (i.e., the data-generating odds ratio differs from one), and thus the Bayes factor B_{10} for the alternative hypothesis increases with sample size. This follows from the fact that the chosen prior distribution ensures consistency, meaning that the Bayes factor for the data-generating model increases as sample size increases (Liang et al., 2008). For a medium-sized effect (OR = 1.67), moderate evidence for the alternative hypothesis can already be obtained with relatively small samples. For instance, for N = 400 observations (which is more than twice the median N of previous studies; cf. Table 1), 48% of the simulated Bayes factors were larger than three ($B_{10} > 3$), and for N = 3,200 and larger, this was the case for all Bayes factors (100%). In contrast, relatively large sample sizes are required to obtain moderate evidence for a small effect (OR = 1.25). As is apparent in the middle panel of Figure A1, only 14% of the simulated Bayes factors were larger than three for a sample size of N = 400, and even for N = 3,200, only 72% of the Bayes factors were larger than three. Note that the median Bayes factor even prefers the (incorrect, but more parsimonious) null hypothesis for sample sizes up to N = 800 (which is close to the maximum sample size of the reanalyzed studies listed in Table 1), since the data do not provide sufficient evidence in favor of a nonzero, but small effect. This shows that uncommonly large sample sizes are required to detect small effects. In the left panel of Figure A1, the true data-generating odds ratio is one (i.e., there is no effect in the population), and thus the Bayes factor shows higher evidence for the null hypothesis as sample size increases. However, in small samples, the Bayes factor is often ambiguous. For instance, with a sample size of N = 400, only 35% of the simulated Bayes factors showed moderate evidence for the null hypothesis (i.e., $B_{01} > 3$).

Overall, these results highlight the importance of having a sufficiently large number of observations to distinguish between small and null effects. If the sample size is relatively small, the simulation showed that the probability is high that the Bayes factor indicates only ambiguous evidence for or against the null hypothesis.