

Managing Self-Presentation: How Social Cues Shape Different Forms of Socially Desirable Responding



Vojin Simunović , Iris Žeželj 

Department of Psychology, Faculty of Philosophy, University of Belgrade, Serbia

Socially desirable responding (SDR) is usually treated as a “noise” in psychological research, to be controlled for by creating certain conditions for respondents. We tested a range of cues aimed to decrease/increase SDR to be applied/avoided in selection or recruitment. To decrease it, we developed two novel procedures: one inspired by the bogus pipeline in which the respondents were induced to believe we can objectively record their sincerity (all three studies), and the other, inspired by the Bayesian Truth Serum (BTS) method, in which the respondents were rewarded points for sincerity (Study 3). To increase it, we exposed the respondents to descriptive group norms signaling socially desirable behaviors in their peers (Study 1 and Study 2). We measured SDR via overclaiming (Study 1 and Study 3), L and K scales from the MMPI (Study 1), and attitudes towards vulnerable groups (Study 2). Across all three studies, we decreased the SDR via newly developed procedures, but failed to increase it, indicating a “default” level of positive self-presentation. When we compared the two procedures for decreasing SDR (overclaiming indices), the one inspired by the bogus pipeline was more effective than the BTS-inspired one (Study 3).

Key words: socially desirable responding, self-presentation, attitudes towards vulnerable groups, bogus pipeline, Bayesian Truth Serum (BTS), overclaiming, descriptive social norms

People are notorious for trying to leave a good impression. Aspects of positive self-presentation (e.g., impression management) may represent an essential human tendency worth exploring per se (Schlenker & Pontari, 2000; Uziel, 2010). On the other hand, socially desirable responding (SDR) can be a threat

to the fairness in employment interviews (Schlenker & Pontari, 2000), and is typically treated as a “noise” in psychological research (Paulhus, 1991, 2003). This led to a number of attempts to detect and/or decrease this tendency. In this study, we contrast interventions aiming to increase and decrease SDR, and

Correspondence concerning this article should be addressed to Iris Žeželj, Faculty of Philosophy, Čika Ljubina Street 18-20, Belgrade, Serbia. E-mail: izezelj@f.bg.ac.rs

Received October 13, 2021



observe their impact on several self-reported measures.

We defined SDR as a tendency towards positive self-presentation in responding to personality inventories and attitude measurement (Paulhus, 1991), i.e., a tendency to provide the responses that make the impression of a person more in accordance with moral norms and social conventions (Paulhus, 2003).

SDR in Psychological Research

Initially, the SDR scales were composed of items referring to the socially desirable behaviors that are not frequent (Hartshorne & May, 1930), as well as to presumably widespread socially undesirable behaviors (e.g., Eysenck & Eysenck, 1964), in order to identify the respondents who are trying to present themselves in a positive manner. There were also attempts to directly assess SDR by asking the respondents to rate the desirability of different behaviors (Edwards, 1957; Jackson & Messick, 1962). Among the most widely used SDR scales are the so-called L and K scales from the MMPI questionnaire (Hathaway & McKinley, 1951; Meehl & Hathaway, 1946).

SDR was treated as a one-dimensional construct until Paulhus (1984; 1991) argued it consists of: 1) self-deception: the tendency to give self-reports that are honest but positively biased and 2) impression management: deliberate self-presentation to an audience. The two-dimensional factorial structure of SDR has since then gathered a lot of empirical support (Uziel, 2010).

Previously described SDR operationalizations were criticized for lacking an objective criterion for determining whether a response is honest or modified by SDR. In an attempt to overcome this, and to operationalize impression management, Paulhus and his associates developed a procedure in which respondents

assessed their familiarity with a number of concepts in 10 different areas; 20% of concepts were nonexistent. They named it the Overclaiming Technique (Paulhus et al., 2003; Paulhus & Harms, 2004), and operationalized SDR as a tendency to claim knowledge of nonexistent items – i.e., to overclaim (Paulhus et al., 2003).

The evidence for validity of the overclaiming measure is mixed. For example, the overclaiming tendencies were positively related to IQ scores, school grades, peer ratings and self-ratings of cognitive ability (Paulhus & Harms, 2004). It was also more prominent in more narcissistic respondents (Ludeke & Makrinski, 2015). On the other hand, overclaiming measure turned out to be poor predictor of measures of job applicants' faking (Feeney & Goffin, 2015), and it is also weakly related to the items from HEXACO personality scale and Self-esteem scale that are rated as socially desirable (Kam et al., 2015). In addition, in a recent study by Bensch et al. (2017), the overclaiming tendency did not load on any of the six first-order factors derived from different SDR and personality measures, and correlations between the overclaiming and other SDR measures varied between .07 and .35. Bensch et al. (2017) concluded that overclaiming is the most distinct potential indicator of positivity bias and that it is independent of known personality measures. However, since overclaiming was constructed with an aim to measure socially desirable responding (Paulhus et al., 2003; Paulhus & Harms, 2004), we decided to treat overclaiming and more standard measures of SDR as manifestations of the same overarching construct.

SDR bias also affects self-reported attitudes, especially attitudes towards vulnerable social groups. It is often named as one of the reasons for attitude-behavior discrepancies. For example, although racism, homophobia and prejudice against women drastically de-

creased in public opinion polls, the number of interracial marriages or friendships and discrimination in the workplace did not follow this change (Hardin & Banaji, 2013; McConnell & Leibold, 2001). Neutralizing or “taming” SDR is one of the often-stated goals of new techniques for implicit measurement of attitudes (for review see Wilson & Scior, 2014; Žeželj et al., 2010). This is why we included biases in attitude reporting as another indicator of SDR tendencies.

Social Cues and Socially Desirable Responding

Decreasing SDR. In spite of the observed individual differences, there is also evidence that SDR tendencies can be situationally influenced. The simplest way to do so is to directly instruct respondents to “fake good” or “fake bad” impression (review in Bou Malham & Saucier, 2016; Tonković et al., 2011; Viswesvaran & Ones, 1999). This turned out to be an effective way to impact respondents’ scores on personality measures (Bagby et al., 1995; Furnham & Henderson, 1982; Graham et al., 1991), and even more their scores on social desirability and lying scales (Viswesvaran & Ones, 1999). There were considerably fewer attempts to affect reporting attitudes using such instructions (Eyssel & Ribas, 2012). These results show that respondents can, at least partially, control their SDR tendencies when directed to do so. However, direct instructions cannot be expected to have an effect if respondents are motivated to present themselves in a certain way (e.g., in a selection process). To this end, different, less obvious situational cues remain to be created and empirically tested.

The researchers used, for example, the so-called “bogus pipeline” (Jones & Sigall, 1971), a device presented to participants as a tool that can objectively assess whether they re-

spond truthfully. This experimental procedure was designed with an aim to convince subjects that a sophisticated new device could accurately detect their “true” attitudes and opinions (Brunell & Buelow, 2019; Larson, 2018; Roese & Jamiseon, 1993; Sassenrath, 2019). The standard procedure typically involves a non-functioning polygraph; participants are attached to a device via electrodes and told it can detect dishonest responses by measuring involuntary responses such as heart rate and skin conductance (Larson, 2018; Roese & Jamiseon, 1993; Suschinsky et al., 2020). The underlying assumption is that respondents would be motivated to respond sincerely, to avoid being caught in a lie (Roese & Jamiseon, 1993; Römer et al., 2018). In this situation, respondents are more likely to provide responses that are not socially desirable – to express negative emotions, report grandiose narcissism and selfishness, smoking and consuming alcohol, and admit negative attitudes towards vulnerable groups (review in: Donahue, 2014; Roese & Jamieson, 1993). In one study of anti-Semitism (Imhoff & Banse, 2009) such manipulation effectively led to an increase in the self-reported anti-Semitic tendencies, and also increased correlations between explicit and implicit measures of attitudes. Similar procedures increased honesty in reporting on political voting (Hanmer et al., 2014) and in reporting on cheating behavior (Fisher & Brunell, 2014). Recent findings show that the bogus pipeline procedure is also effective in decreasing SDR in the context of self-reporting on a diverse set of outcomes such as: well-being and narcissism (Brunell & Buelow, 2019), intimate desires (Suschinsky et al., 2020), empathic responses (Sassenrath, 2019), and income information (Römer et al., 2018).

Another strategy for decreasing SDR is directly incentivizing truth telling and penalizing cheating/lying: for example, there is a sur-

vey scoring system called the Bayesian Truth Serum (BTS), that rewards honesty or certain type of responses to multiple-choice questions (John et al., 2012; Prelec, 2004; Weaver & Prelec, 2013). Inspired by the concept of Nash-equilibrium from the game theory (Myerson, 1999), the key idea behind BTS is to “assign a high score to an answer whose actual frequency is greater than its predicted frequency, with predictions drawn from the same population that supplies the answers” (Weaver & Prelec, 2013, p. 1). BTS proved to effectively discourage the recognition of fictitious concepts on the overclaiming questionnaire (Weaver & Prelec, 2013). In the same research, the question of the effectiveness of different ways to decrease SDR was examined, and BTS outperformed the “solemn oath”, a rival truth-inducing mechanism, while the bogus pipeline procedure was mentioned but not included in the experiments (Weaver & Prelec, 2013).

Increasing SDR. Although decreasing SDR is far more explored due to its more obvious potential for implementation, we argue that, to know more about the nature of the construct, we need to explore whether it can be experimentally increased, as well. Comparing a control group with a group induced to SDR could inform us about the “default” level of socially desirable responding. In addition, the enhancing procedures, if successful, might provide practitioners with valuable information on what to avoid. Apart from already discussed “fake good” instructions, there were very few empirical attempts to experimentally increase SDR. We thought there are experimental techniques that could be borrowed from other research fields to this end. For example, one can draw from a vast literature on normative influences on behavior which shows that, when presented with a descriptive norm (i.e., how other people behave, feel or think) people tend to adjust their behav-

ior to fit it more (Cialdini, 2007; McDonald & Crandall, 2015; Rimal & Real, 2003). Following that logic, if we expose the participants to group norms stating that the majority of their peers behave in a socially desirable way, we can expect they would adjust their answers to fit the descriptive norm.

Overview of Present Research

We designed two experiments in which we tried to both enhance and decrease SDR and compare it to the control group. To decrease SDR, we used an experimental manipulation similar to the bogus pipeline, and to increase it, we exposed the participants to descriptive peer group norms. To test the relation between the constructs and the generalizability of the effects, we measured SDR in three domains: self-image (scales L and K from the MMPI questionnaire), general knowledge (overclaiming technique) and attitudes towards vulnerable groups. In the third study, we compared the effectiveness of two experimental procedures aimed at decreasing the scores on overclaiming indices: one inspired by the bogus pipeline and the other inspired by the BTS method.

These studies add to the existing literature in several important ways: a) they test whether SDR is sensitive to situational cues, i.e., whether the respondent’s awareness of situational cues can influence this type of responding, b) they contrast two interventions aimed to decrease and increase SDR in a single design, as well as compare the effects of two different decreasing procedures, and finally, c) they relate different types of SDR and compare how malleable they are.

Study 1

The main aim of this study was to explore whether it is possible to reduce or increase

SDR measured by the L and K scales from the MMPI questionnaire, as well as by the overclaiming questionnaire. These measures were selected because they are designed to measure two different dimensions of SDR: self-deception and impression management. In addition, we wanted to explore relations between overclaiming and the L and K scales, as alternative measures of socially desirable responding. Although the bogus pipeline is a widely used technique for encouraging honest responding, its effect on these types of SDR has, to our knowledge, not been tested. We developed two novel procedures to experimentally decrease/increase SDR. The first was a logistically less demanding version of the bogus-pipeline, and the second was based on social comparison.

Method

Hypotheses

We expected that the respondents convinced that the truthfulness of their responses is objectively assessed will manifest a lower tendency towards positive self-presentation than the control group, whilst the respondents exposed to descriptive group norms will manifest a higher tendency towards positive self-presentation than the control group (H1).

All three dependent variables were expected to correlate positively; this would point to one latent tendency towards positive self-presentation, which has more modalities (H2).

The study design and hypotheses were uploaded to Moodle online platform prior to data collection, as a part of a Master's Thesis submission. The research was approved by the IRB of the Department of Psychology (Protocol numbers: 2020-48 and 2022-23).

All data, instruments, syntax and power curves are available at the OSF platform (<https://osf.io/yswdh/>).

Instruments

We used Paulhus's overclaiming technique adapted for Serbian respondents (Kašiković et al., 2013). In that technique, SDR is operationalized through the accuracy index and the bias index (Paulhus et al., 2003). This version contains 150 concepts (e.g., existent: The Boston Tea Party, behaviorism; nonexistent: otoplasm, Gambian peace), with 20% of them being nonexistent (Kašiković et al., 2013). Respondents rated their familiarity with the listed concepts on the scale from 1 (*not familiar at all*) to 4 (*fully familiar*). Bias (overclaiming) and accuracy (knowledge of concepts) indices were calculated following the Signal Detection Theory, including 4 response categories: recognizing existing items (hits), recognizing nonexistent items (false alarms), not recognizing existing items, and not recognizing nonexistent items (Paulhus et al., 2003). The bias index was calculated using the formula: $\text{Bias} = [\text{Hits} + \text{False alarms}]/2$. The accuracy index was calculated using the formula: $\text{Accuracy} = \text{Hits} - \text{False alarms}$ (Paulhus et al., 2003).

We also used scales L and K from the MMPI-202 (Biro, 2008), the latest Serbian version of the MMPI questionnaire, that is widely used in clinical testing in Serbia.

L scale (lying scale) from the standardized 2008 Serbian version of the MMPI questionnaire (Biro, 2008), contains 14 statements (e.g., "Sometimes I lie"), aimed to measure extreme self-promotion. The respondents answer with "yes" or "no"; summary score can vary from 0 to 14. In clinical testing, respondent's result on the MMPI questionnaire is considered valid if the score on the L scale is between 0 and 8 points (for both genders); The 50th percentile is approximately 4 points (Biro, 2008).

K scale (correction scale) from the standardized 2008 Serbian version of the MMPI

questionnaire (Biro, 2008), contains 26 yes/no statements (e.g., "Criticism and reproach strongly affect me") aimed to measure subtle self-promotion. Total score varies from 0 to 26. In clinical testing, respondent's result on the MMPPI questionnaire is considered valid if the score on the K scale is between 2 and 21 (for both genders); the 50th percentile is approximately 12 points (Biro, 2008).

Procedure

We tried to increase positive self-presentation by exposing respondents to false information about the percent of other respondents (their colleagues), who reported knowledge of the presented concepts (for overclaiming) or gave responses that bring points on the L or K scale. Each concept and each question were presented on a separate screen. The fictitious percent varied from 50% to 90% because we expected that 50+ percent will stimulate the effect of above average self-evaluation (Alicke, 1985) on the overclaiming questionnaire, as well as social comparison on the MMPPI questionnaire.

We tried to decrease positive self-presentation by asking the respondents to verbalize their answers using a microphone that was connected to a computer. They were told that their voice is analyzed by a new lie detection software that detects voice modulations; this is an adapted version of a typical bogus pipeline procedure (Imhoff & Banse, 2009).

Respondents from the control group did not receive any instructions other than standard ones related to the content of the tasks.

All respondents were individually tested; they were projected the questions one by one and asked to read them out loud, and verbally respond to them, while the examiner was writing it down. They were informed that their answers will be joined together in a database and anonymized for further analyses.

In an exit interview, we probed for suspicion of our experimental manipulation (we did not have to exclude anyone based on this criteria) and fully debriefed the respondents.

Respondents

A total of 52 first-year psychology students (48 females) took part in exchange for course credit. They were randomly assigned to experimental groups. 18 participants were induced to sincerity, 18 were induced to SDR, and 16 were in the control group.

Results

Generally, our respondents did not manifest a strong tendency towards positive self-presentation (Table 1). The average scores on the L and K scales were close to the population average, suggesting that subtle as well as strong self-presentation strategies within our respondents were moderate. Judging by their scores on the overclaiming questionnaire, our respondents also manifested a modest tendency to overestimate their own general knowledge.

The Effects of Social Cues on Different Forms of SDR

To test whether the experimental induction was successful, we employed one-way between-subjects ANOVA analyses with scores on scales L and K and overclaiming indices as dependent variables. The experimental groups differed on both MMPPI scales. The overall effect on the L scale was significant ($F(2, 49) = 13.41$; $p < .001$; $\eta^2 = .35$); however, Tukey post hoc test revealed that the difference between the group induced to sincerity and the control group was significant ($HSD = 2.06$; $p < .001$; $d = 1.34$), and the difference between the group exposed to descriptive norms and

Table 1 *Positive self-presentation tendencies in three experimental groups*

	L scale (0-14)	K scale (0-26)	Bias index (75-300)	Accuracy index (0-450)
Group induced to SDR	4.78/0.5 (1.35/0.76)	13.68/0.54 (3.45/0.95)	196.97/0.09 (20.58/0.82)	304.39/0.06 (33.04/0.77)
Control group	4.5/0.34 (1.67/0.94)	11.75/0.01 (3.01/0.83)	205.28/0.34 (27.57/0.93)	314.31/0.38 (37.7/1.04)
Group induced to sincerity ("Lie Detection Software")	2.44/-0.8 (1.38/0.77)	9.72/-0.55 (3.39/0.93)	182.65/-0.39 (28.05/1.13)	284.89/-0.4 (45.74/1.07)
Total sample	3.88/0 (1.79/1)	11.71/0 (3.64/1)	194.8/0 (26.67/1)	300.69/0 (40.38/1)

Note. We report average scores/standardized average scores with standard deviations/standardized standard deviations in brackets.

Table 2 *Correlations between different positive self-presentation measures*

	1	2	3	4
L scale		.43 <u>$p = .002$</u>	.28 <u>$p = .051$</u>	.18 $p = .192$
K scale	-		.13 $p = .35$.14 $p = .339$
Bias index	-	-		.93 <u>$p < .001$</u>
Accuracy index	-	-	-	

the control group was not ($p = .584$). Similarly, the overall effect was also significant on the K scale ($F(2, 49) = 6.42$; $p = .003$; $\text{Eta}^2 = .21$), but Tukey post hoc test showed that the difference between the group induced to sincerity and the control group ($p = .185$), as well as between the group exposed to descriptive norms and the control group ($p = .22$) was not. The difference between groups on the bias index was marginally significant¹

¹ The term marginally significant can sometimes be incorrectly used to suggest that the effect is "approaching" significance of .05. We used the term in this paper simply to acknowledge p coefficient between .05 and .1.

($F(2, 49) = 2.8$; $p = .071$; $\text{Eta}^2 = .1$); the difference between the group induced to sincerity and the control group was marginally significant (HSD = 29.42; $p = .084$; $d = 0.7$), and the difference between the group exposed to descriptive norms and the control group was not ($p = .743$). The difference between groups on the accuracy index was also marginally significant ($F(2, 49) = 2.5$; $p = .092$; $\text{Eta}^2 = .09$). Once again, the difference between the group induced to sincerity and the control group was marginally significant (HSD = 20.67; $p = .06$; $d = 0.74$) and the difference between the group exposed to descriptive norms and the

control group was not ($p = .617$). The effects of the experimental manipulation were significant for all measures, and Cohen's d s for the differences between the group induced to sincerity and the control group were moderate or high.

Correlation between SDR Measures

Correlation between L and K scales was relatively high, while correlation between L scale and bias index was moderate and marginally significant. Extremely high correlation between accuracy index and bias index is an artifact of the fact that their formulas are based on the same parameters (Table 2).

Study 2

In Study 1 we examined the impact of experimental manipulation on the scores on positive self-presentation in reporting behavior and in recognizing concepts from the general culture. But SDR is also being manifested in other domains, for example, in attitudes towards different groups. To determine whether the variation of the situation factors can affect this form of SDR, we have designed Study 2. As psychologists opted for the so-called "helping profession" and are typically ideologically liberal (Graziano & Habashi, 2010; Imhoff & Banse, 2009; Swank & Raiz, 2010), they make a particularly interesting group for examining the impact of experimental manipulation on attitudes towards vulnerable groups. The underlying idea of the design is similar to the idea of implicit measurement of attitudes (Žeželj et al., 2010; Wilson & Scior, 2014) – to create experimental conditions that would discourage socially desirable responding, i.e., expressing tolerance and non-discriminatory attitudes towards the Roma people, homosexuals, users of psychiatric services, narcotic users, and children with disabilities.

Method

Instrument

The dependent variable was the score on the questionnaire of attitudes towards five vulnerable groups (Roma people, homosexuals, users of psychiatric services, narcotic users and children with disabilities). Attitude towards each group was represented by five statements; every statement was evaluated on a Likert-type scale anchored with 1 (*totally disagree*) to 4 (*totally agree*). To create a discriminative measure, we worded the statements (e.g., "Drug addiction is a reflection of weak character") to reflect a range of attitudes, having in mind the targeted respondents whom we expected to be more tolerant than the general population. The general scale of attitudes towards vulnerable groups had $\alpha = .83$ level of reliability, which allowed us to compute an average total score. The reliability of specific scales of attitudes varied from $\alpha = .445$ (Roma people) to $\alpha = .821$ (homosexuals) with four out of five higher than $\alpha = .5$, so we also calculated five specific scores.

Procedure

We applied the same experimental procedure detailed in Study 1.

Respondents

Based on the effect size from the first study ($\text{Eta}^2 = .35$ on the L scale), we calculated using G*power 3 software (Faul et al., 2007) that a total of 33 respondents would be enough to reach .95 level of statistical power. However, as 51 psychology students (39 females) volunteered to take part, we ended up testing all of them; there were 17 respondents randomly assigned to each group.

Results

Generally, our respondents manifested tolerant attitudes towards vulnerable groups, however there was still room for increase – i.e., there was no ceiling effect (detailed in Table 3).

The Effects of Social Cues on Attitudes towards Vulnerable Groups

To test whether the experimental induction was successful, we employed one-way between-subjects ANOVA analyses, with scores on general scale as well as scores on each scale of attitudes towards vulnerable groups as dependent variables. As expected, we observed differences between the experimental groups on the general attitude measure ($F(2, 48) = 8.71; p = .001; \text{Eta}^2 = .27$), and Tuckey post hoc test indicated that the difference between the group induced to sincerity and the control group was significant (HSD = 0.32; $p = .035; d = 0.88$), and the difference between the group exposed to descriptive norms and the control group was not ($p = .27$). As for tolerance towards specific vulnerable groups, we observed a difference between groups on the attitudes towards *homosexuals* ($F(2, 48) = 6.45; p = .003; \text{Eta}^2 = .21$); the difference between the group induced to sincerity and the control group was significant (HSD = 0.67; $p = .022; d = 0.92$), and the difference between the group exposed to descriptive norms and the control group was not ($p = .806$). The difference between groups on the attitudes towards *Roma people* was marginally significant ($F(2, 48) = 2.91; p = .064; \text{Eta}^2 = .11$); while the difference between the group induced to sincerity and the control group ($p = .475$), as well as between the group exposed to descriptive norms and the control group ($p = .435$) was not. Finally, there were differences between groups on

the attitude towards *children with disabilities* ($F(2, 48) = 7.22; p = .002; \text{Eta}^2 = .23$); this was the only case in which the difference between the group exposed to descriptive norms and the control group was significant (HSD = 0.36, $p = .05; d = 0.83$) and the difference between the group induced to sincerity and the control group was not ($p = .387$).

The effects of the experimental manipulation were significant on three out of five scales as well as on the general measure, and Cohen's d s for the differences between groups were moderate or high.

Study 3

The results of Study 1 and Study 2 showed that it is easier to decrease than to increase the level of SDR. To further put our newly developed procedure to the test, we conducted an additional study, in which we a) replicated the results in a larger sample, and b) contrasted this manipulation with another type of strategy for decreasing SDR – direct incentivization.

Instrument

We used the same instrument as in Study 1 – Paulhus's overclaiming technique adapted for Serbian respondents (Kašiković et al., 2013); the overclaiming tendency is operationalized through the accuracy index and the bias index (Paulhus et al., 2003). Due to the difference in the procedure ("Quiz" situation), this time, instead of a four-point scale, the respondents were asked if they had heard about each concept, and they answered with "yes" or "no". The formulas for calculating the overclaiming indices were the same as in Study 1.

Procedure

We applied similar experimental procedures as in Study 1 and Study 2.

Table 3 Attitudes towards vulnerable groups in three experimental groups

	Roma people (1-4)	Homosexuals (1-4)	Users of psychiatric services (1-4)	Narcotic users (1-4)	Children with disabilities (1-4)	Average attitude towards vulnerable groups (1-4)
Group induced to SDR	3.18/0.4 (0.48/0.93)	3.12/0.41 (0.69/0.88)	3.18/0.24 (0.48/0.94)	3.22/0.35 (0.55/0.88)	3.18/0.63 (0.45/0.92)	3.17/0.57 (0.36/0.86)
Control group	2.96/-0.08 (0.45/0.88)	2.96/0.22 (0.71/0.9)	3.14/0.18 (0.43/0.83)	3.01/0.01 (0.66/1.05)	2.81/-0.11 (0.44/0.9)	2.98/0.1 (0.42/1.01)
Group induced to sincerity ("Lie Detection Software")	2.76/-0.39 (0.55/1.07)	2.29/-0.63 (0.74/0.94)	2.8/-0.42 (0.58/1.12)	2.79/-0.35 (0.62/0.99)	2.6/-0.52 (0.42/0.86)	2.66/-0.67 (0.3/0.73)
Total sample	2.97/0 (0.52/1)	2.79/0 (0.78/1)	2.87/0 (0.49/1)	2.97/0 (0.52/1)	2.87/0 (0.49/1)	2.94/0 (0.42/1)

Note. We report average scores/standardized average scores with standard deviations/standardized standard deviations in brackets.

One experimental manipulation, aimed to decrease SDR, was the same as in previous studies: we asked the respondents from this group to verbalize their answers using a microphone that was connected to a computer, and told them that their voice is analyzed by a new lie detection software which detects voice modulations; this is an adapted version of a typical bogus pipeline procedure (Imhoff & Banse, 2009; Larson, 2018; Römer et al., 2018).

Another experimental way to decrease SDR was to incentivize truthful responding; this approach is inspired by the BTS procedure (John et al., 2012; Prelec, 2004). Respondents from this group were introduced into the "Quiz" situation and were told they will get one quiz point for each existing concept they recognize as familiar, as well as for a fictitious concept they say is unfamiliar to them; but that they will get a negative point for every fictitious concept they said they are familiar with. After each response, they could see on the screen whether the answer was correct and whether they would gain or lose a point.

Respondents from the control group did not receive any instructions other than the standard ones related to the content of the tasks.

Respondents

A total of 90 students of the Faculty of Philosophy (69 females) took part in our study. They were randomly assigned to three experimental groups: 1) attached to "Lie Detection Software", 2) "truth telling incentivized", and 3) control group. There were 30 respondents in each group.

Results

Judging by the average scores on the overclaiming indices, our respondents moderately overestimated their own general knowledge (Table 4).

The Effects of Social Cues on Scores on Overclaiming Indices

To test whether the experimental induction was successful, we employed one-way be-

Table 4 *Overclaiming indices in three experimental groups*

	Bias index (range: 0-75)	Accuracy index (range: -25 - 125)
"Lie Detection Software" group	41.85/-0.67 (6.27/0.81)	72.03/-0.49 (9.54/0.76)
"Truth telling incentivized" group	47.73/0.09 (7.36/0.95)	78.73/0.04 (14.66/1.17)
Control group	51.55/0.58 (6.51/0.84)	83.77/0.45 (10.18/0.81)
Total sample	47.04/0 (7.77/1)	78.18/0 (12.52/1)

Note. We report average scores/standardized average scores with standard deviations/standardized standard deviations in brackets.

tween-subjects ANOVA analyses, with scores on two overclaiming indices as dependent variables. The groups differed on the bias index ($F(2, 89) = 15.82; p < .001; \eta^2 = .267$) – the Tukey post hoc test revealed that the group attached to “Lie Detection Software” had lower scores than the group in the “Quiz” situation (HSD = 5.88; $p = .003; d = 0.86$) and the control group (HSD = 9.7; $p < .001; d = 1.52$). The difference between the group in the “Quiz” situation and the control group was marginally significant (HSD = 3.82; $p = .077; d = 0.55$).

The experimental groups also differed on the accuracy index ($F(2, 89) = 7.61; p = .001; \eta^2 = .149$). Once again, the Tukey post hoc test revealed that the difference between the group attached to “Lie Detection Software” and the control group was significant (HSD = 11.73; $p = .001; d = 1.19$), and the difference between the group attached to “Lie Detection Software” and the group in the “Quiz” situation was marginally significant (HSD = 6.7; $p = .073; d = 0.54$), whilst the difference between the group in the “Quiz” situation and the control group was not ($p = .223$).

Both overclaiming indices were significantly affected by the experimental manipulation, and Cohen’s d s for the differences between groups were moderate or high.

Discussion

Across all three studies, we managed to *decrease* SDR among psychology students by introducing a lie detection software allegedly sensitive to voice modulations. In Study 1, the respondents in this condition were less prone to use strong self-presentation strategies measured by the L scale from the MMPI questionnaire, as well as less prone to claim they were familiar with fictitious concepts, measured by the overclaiming questionnaire. These results are consistent with previous

findings (Larson, 2018; Roesse & Jamiseon, 1993; Römer et al., 2018; Suschinsky et al., 2020), showing the effectiveness of the bogus pipeline procedure in reducing SDR in different domains.

On the other hand, “lie detection software” was not as efficient in decreasing subtle self-presentation measured by the K scale, which is probably due to the fact that the K scale measures self-deception, which is less controllable than impression management (Uziel, 2010). In Study 2, respondents attached to the “lie detector” were consistently reporting less tolerant attitudes towards different vulnerable groups. We aimed to *increase* SDR by making it a norm (showing a fictitious majority of peers who responded in a socially desirable manner): in neither of the two studies was this group different from the control.

There can be several potential reasons for this asymmetry: one can say that they might be due to the restriction of range, i.e., that the respondents were initially extreme in their answers so they could be decreased but not further increased. However, this remark is relevant only to Study 2. In this study the respondents in the control group, which is informative of the baseline attitudes, were tolerant towards vulnerable groups (but still not extreme, however). In Study 1, the respondents were moderate in each of the self-presentation strategies, nevertheless, the manipulation led only to their decrease not increase. Additionally, one might say that these two levels of experimental manipulation do not mirror one another, so it might just be the case that the bogus lie detector is simply more effective than evoking descriptive norms. While there is certainly merit to this view, we opted for the two levels of manipulation after careful deliberation. We did not find the standard fake good/fake bad procedure suitable as it, in our view, tests respondents’ knowledge on how to answer in order to present themselves

in a certain way – in other words, a meta cognitive ability of a kind, whilst we aimed to create situations in which they will be genuinely motivated to present themselves in a certain manner. It was not possible to create a direct counterpart to the fictitious lie detector situation so as to motivate respondents to lie/positively self-present (it would be very implausible if, for example, we told them that the software aims to encourage respondents to do so). Having in mind strong empirical evidence to the power of descriptive norms in shaping the behavior (Cialdini, 2007), we consider the manipulation we choose to be both plausible and ecologically valid. Hence, although we do not claim the two procedures were directly comparable regarding their effect sizes, we argue there is enough evidence to suggest a stable level of “default” self-presentation that is difficult to further increase by situational cues. Results from similar studies in which the researchers even used direct instructions (e.g., Viswesvaran & Ones, 1999) also corroborate our conclusion that it is easier to decrease than to increase self-presentation tendency.

Relationship between SDR Measures

Despite apparent differences in conceptualization and methodical design, we observed positive correlation between the L scale and the bias index on overclaiming; this correlation was higher than the one reported in Bensch et al. (2017), indicating these measures have a common core. Both measures are related to impression management, and that core may be a certain tendency to present oneself as a well-educated and conventionally socialized individual.

On the other hand, we did not observe a correlation between the K scale and the bias index, suggesting there are important differences in the psychological mechanisms behind the measures. That is in line with the

fact that the K scale measures self-deception, while the overclaiming technique is aimed to operationalize impression management (Bensch et al., 2017).

Correlation between scales L and K can be partly attributed to the shared methodological variance; however, it can also indicate the existence of similar underlying tendencies. While we observed positive relations between these constructs, the study design and sample size do not allow us to give any definite answers regarding their nature.

Comparison between the “Truth telling incentivized” and the “Lie Detection Software” Procedures

Following up the question of the effectiveness of different ways to reduce SDR (Weaver & Prelec, 2013), we examined (in Study 3) whether the bogus pipeline or the BTS procedure is more effective. The results presented in Table 4 showed that the group examined in the bogus pipeline procedure had lower scores on both overclaiming indices than the group examined in the BTS-inspired procedure, which is a result that indicates that the bogus pipeline is a more effective way to decrease SDR. This finding could also indicate that avoiding being “caught in a lie” is a stronger motivator for truthful responding than reinforcement in the form of points on a quiz.

Limitations and Future Research

Since the samples in the first two studies were relatively small, future studies are warranted to replicate them on a larger sample.

Also, our samples were unbalanced by gender (especially in Study 1), due to the fact that the vast majority of psychology students are females. Hence, we would recommend using a sample with similar proportions of female and male respondents.

We opted for between-subjects design so that the respondents were less likely to be suspicious of the manipulation. Ideally, to account for noise originating in individual differences, a pretest-posttest design with a long period in between should be employed and the results compared to the ones we obtained.

We can also suggest an alternative manipulation for stimulating SDR: for overclaiming, and, to a lesser extent, L and K scores – the experimental situation could simulate job or other type of selection. Alternatively, the respondents could be informed that the test score correlates with some central abilities, such as IQ score. For tolerance towards vulnerable groups, future researchers could introduce a member of a particular group to the respondent in an experimental situation thus facilitating socially desirable responding. The proposed technique, however, would be challenging to implement, and it would require pretesting the prototypicality of a particular group member.

Finally, the “Truth telling incentivized” group in Study 3 was rewarded with points on a quiz, rather than with money as in earlier research (John et al., 2012; Prelec, 2004; Weaver & Prelec, 2013); that could have affected the results, so future research should vary the reward for respondents in that group.

Although our findings do not give a definite answer to the question whether it is possible to experimentally increase SDR, they consistently testify to the fact that an easy to implement intervention can decrease it. Keeping in mind that if we need to inform policy makers about policies sensitive to socially desirable bias, decreasing it would be vital and this research provides empirical support for a relatively simple way to do so.

Acknowledgment

This research is prepared as a part of the project No. 451-03-47/2023-01/ 200163, funded

by Ministry of Education, Science and Technological Development of Republic of Serbia.

Authors' ORCID

Vojin Simunović

<http://orcid.org/0000-0001-6526-4208>

Iris Žeželj

<http://orcid.org/0000-0002-9527-1406>

References

- Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, *49*, 1621–1630. <https://doi.org/10.1037/0022-3514.49.6.1621>
- Bagby, R. M., Buis, T., & Nicholson, R. A. (1995). Relative effectiveness of the standard validity scales in detecting fake-bad and fake-good responding: Replication and extension. *Psychological Assessment*, *7*, 84. <http://doi.org/10.1037/1040-3590.7.1.84>
- Bensch, D., Paulhus, D. L., Stankov, L., & Ziegler, M. (2017). Teasing apart overclaiming, overconfidence, and socially desirable responding. *Assessment*, *26*, 351–363. <https://doi.org/10.1177/1073191117700268>
- Biro, M. (2008). *Dijagnostička procena ličnosti – MMPI-202 [Diagnostic Assessment of Personality - MMPI-202]*. Beograd: Centar za primenjenu psihologiju.
- Bou Malham, P., & Saucier, G. (2016). The conceptual link between social desirability and cultural normativity. *International Journal of Psychology*, *51*, 474–480. <https://doi.org/10.1002/ijop.12261>
- Brunell, A. B., & Buelow, M. T. (2019). Using the bogus pipeline to investigate trait narcissism and well-being. *Personality and Individual Differences*, *151*, 109509. <https://doi.org/10.1016/j.paid.2019.109509>
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *The Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis, MN: University of Minnesota Press.
- Cialdini, R. B. (2007). Descriptive social norms as underappreciated sources of social control. *Psychometrika*, *72*, 263–268. <https://doi.org/10.1007/s11336-006-1560-6>

- Donahue, M. (2014). *On being second guessed by a machine: A reevaluation of the bogus pipeline* (Doctoral dissertation, Wesleyan University).
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.
- Eysenck, F., & Ribas, X. (2012). How to be good (or bad): On the fakeability of dehumanization and prejudice against outgroups. *Group Processes & Intergroup Relations*, 15, 804–812. <http://doi.org/10.1177/1368430212447204>
- Eysenck, H. J., & Eysenck, S. B. G. (1964). *The manual of the Eysenck Personality Inventory*. London: University of London Press.
- Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A revised version of the Psychoticism Scale. *Personality and Individual Differences*, 6, 21–29. [https://doi.org/10.1016/0191-8869\(85\)90026-1](https://doi.org/10.1016/0191-8869(85)90026-1)
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. <https://doi.org/10.3758/BF03193146>
- Feeney, J. R., & Goffin, R. D. (2015). The Overclaiming Questionnaire: A good way to measure faking? *Personality and Individual Differences*, 82, 248–252. <https://doi.org/10.1016/j.paid.2015.03.038>
- Fisher, T., & Brunell, A. B. (2014). A bogus pipeline approach to studying gender differences in cheating behavior. *Personality and Individual Differences*, 61-62, 91–96. <https://doi.org/10.1016/j.paid.2014.01.019>
- Furnham, A., & Henderson, M. (1982). The good, the bad and the mad: Response bias in self-report measures. *Personality and Individual Differences*, 3, 311–320. [http://doi.org/10.1016/0191-8869\(82\)90051-4](http://doi.org/10.1016/0191-8869(82)90051-4)
- Graham, J. R., Watts, D., & Timbrook, R. E. (1991). Detecting fake-good and fake-bad MMPI-2 profiles. *Journal of Personality Assessment*, 5, 264–277. https://doi.org/10.1207/s15327752jpa5702_6
- Graziano, W. G., & Habashi, M. M. (2010). Motivational processes underlying both prejudice and helping. *Personality and Social Psychology Review*, 14, 313–331. <https://doi.org/10.1177/1088868310361239>
- Hanmer, M. J., Banks, A. J., & White, I. K. (2014). Experiments to reduce the over-reporting of voting: A pipeline to the truth. *Political Analysis*, 22, 130–141. <https://doi.org/10.1093/pan/mpt027>
- Hardin, C. D., & Banaji, M. R. (2013). The nature of implicit prejudice: Implications for personal and public policy. In E. Shafir (Ed.), *The behavioral foundations of public policy* (pp. 13–31). Princeton University Press.
- Hartshorne, H., & May, H. A. (1930). *Studies in the nature of character*. New York: MacMillan.
- Hathaway, S. R., & McKinley, J. C. (1951). *MMPI manual*. New York: Psychological Corporation.
- Imhoff, R., & Banse, R. (2009). Ongoing victim suffering increases prejudice. *Psychological Science*, 20, 1443–1447. <https://doi.org/10.1111/j.1467-9280.2009.02457.x>
- Jackson, D. N., & Messick, S. (1962). Response styles on the MMPI: Comparison of clinical and normal samples. *Journal of Abnormal and Social Psychology*, 65, 285–299. <http://doi.org/10.1037/h0045340>
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23, 524–532. <https://doi.org/10.1177/0956797611430953>
- Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, 76, 349–364. <http://doi.org/10.1037/h0031617>
- Kam, C., Risavy, S. D., & Perunovic, W. Q. E. (2015). Using over-claiming technique to probe social desirability ratings of personality items: A validity examination. *Personality and Individual Differences*, 74, 177–181. <http://doi.org/10.1016/j.paid.2014.10.017>
- Kašiković, B., Jerotijević, I., Novaković, S., Mijatović, K., & Žeželj, I. (2013). Kako izmeriti socijalno poželjno odgovaranje: Adaptacija Paulusovog testa “Precenjivanja znanja” [How to measure socially desirable responding: Adaptation of Paulhus “Overclaiming technique”]. Paper presented at 19th meeting Empirical research in psychology, Belgrade, Serbia, book of abstracts pp. 96–97.
- Larson, R. B. (2019). Controlling social desirability bias. *International Journal of Market Research*, 61(5), 534–547. <https://doi.org/10.1177/1470785318805305>
- Ludeke, S., & Makrinski, G. (2015). Does the Over-Claiming Questionnaire measure overclaim-

- ing? Absent convergent validity in a large community sample. *Psychological Assessment*, 28, 765–774. <http://doi.org/10.1037/pas0000211>
- McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, 37, 435–442. <http://doi.org/10.1006/jesp.2000.1470>
- McDonald, R. I., & Crandall, C. S. (2015). Social norms and social influence. *Current Opinion in Behavioral Sciences*, 3, 147–151. <https://doi.org/10.1016/j.cobeha.2015.04.006>
- Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology*, 30, 525–564.
- Myerson, R. B. (1999). Nash equilibrium and the history of economic theory. *Journal of Economic Literature*, 37, 1067–1082. <https://doi.org/10.32609/0042-8736-2010-6-26-43>
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598–609.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. Shaver & L. S. Wrightsman (Eds.), *Measures of Personality and Social Psychological Attitudes*, 17–59. San Diego: Academic Press. <https://doi.org/10.1016/B978-0-12-590241-0.50006-X>
- Paulhus, D. L. (2003). Self-presentation measurement. In R. Fernandez-Ballesteros (Ed.), *Encyclopedia of Psychological Assessment* (pp. 858–861). Thousand Oaks, C. A.: Sage.
- Paulhus, D. L., & Harms, P. D. (2004). Measuring cognitive ability with the overclaiming technique. *Intelligence*, 32, 297–314. <https://doi.org/10.1016/j.intell.2004.02.001>
- Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lycin, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology*, 84, 890–904. <http://doi.org/10.1037/0022-3514.84.4.890>
- Prelec, D. (2004). A Bayesian truth serum for subjective data. *Science*, 306, 462–466. <https://doi.org/10.1126/science.1102081>
- Rimal, R. N., & Real, K. (2003). Understanding the influence of perceived norms on behaviors. *Communication Theory*, 13, 184–203. <https://doi.org/10.1111/j.1468-2885.2003.tb00288.x>
- Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin*, 114, 363–375. <http://doi.org/10.1037/0033-2909.114.2.363>
- Römer, U., Mußhoff, O., Weber, R., & Turvey, C. G. (2018). Assessing the reliability of self-reported income information in informal small business lending through a bogus pipeline experiment. *Journal of Agricultural Economics*, 69, 726–738. <https://doi.org/10.1111/1477-9552.12263>
- Sassenrath, C. (2020). “Let me show you how nice I am”: Impression management as bias in empathic responses. *Social Psychological and Personality Science*, 11(6), 752–760. <https://doi.org/10.1177/1948550619884566>
- Schlenker, B. R., & Pontari, B. A. (2000). The strategic control of information: Impression management and self-presentation in daily life. In A. Tesser, R. B. Felson, & J. M. Suls (Eds.), *Psychological perspectives on self and identity*, 199–232. American Psychological Association. <https://doi.org/10.1037/10357-008>
- Suschinsky, K. D., Fisher, T. D., Maunder, L., Hollenstein, T., & Chivers, M. L. (2020). Use of the bogus pipeline increases sexual concordance in women but not men. *Archives of Sexual Behavior*, 49, 1517–1532. <https://doi.org/10.1007/s10508-020-01737-4>
- Swank, E., & Raiz, L. (2010). Attitudes toward gays and lesbians among undergraduate social work students. *Affilia*, 25, 19–29. <https://doi.org/10.1177/0886109909356058>
- Tonković, M., Galić, Z., & Jerneić, Ž. (2011). The construct validity of over-claiming as a measure of egoistic enhancement. *Review of Psychology*, 18, 11–19.
- Uziel, L. (2010) Rethinking social desirability scales: From impression management to interpersonally oriented self-control. *Perspectives on Psychological Science*, 5, 243–262. <https://doi.org/10.1177/1745691610369465>
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59, 197–210. <https://doi.org/10.1177/00131649921969802>
- Weaver, R., & Prelec, D. (2013). Creating truth-telling incentives with the Bayesian truth serum. *Journal of Marketing Research*, 50, 289–302. <https://doi.org/10.1509/jmr.09.0039>

-
- Wilson, M. C., & Scior, K. (2014). Attitudes towards individuals with disabilities as measured by the Implicit Association Test: A literature review. *Research in Developmental Disabilities, 35*, 294–321. <https://doi.org/10.1016/j.ridd.2013.11.003>
- Žeželj, I., Lazarević, Lj., & Pavlović, M. (2010). Test Implicitnih Asocijacija (IAT): Teorijske i metodološke osnove [The Implicit Association Test: Theoretical and methodological background]. *Psihologijske Teme, 19*, 45–70.