

# A stochastic lie detector

Morten Moshagen · Jochen Musch · Edgar Erdfelder

Published online: 20 August 2011  
© Psychonomic Society, Inc. 2011

**Abstract** Surveys on sensitive issues provide distorted prevalence estimates when participants fail to respond truthfully. The randomized-response technique (RRT) encourages more honest responding by adding random noise to responses, thereby removing any direct link between a participant's response and his or her true status with regard to a sensitive attribute. However, in spite of the increased confidentiality, some respondents still refuse to disclose sensitive attitudes or behaviors. To remedy this problem, we propose an extension of Mangat's (*Journal of the Royal Statistical Society: Series B*, 56, 93–95, 1994) variant of the RRT that allows for determining whether participants respond truthfully. This method offers the genuine advantage of providing undistorted prevalence estimates for sensitive attributes even if respondents fail to respond truthfully. We show how to implement the method using both closed-form equations and easily accessible free software for multinomial processing tree models. Moreover, we report the results of two survey experiments that provide evidence for the validity of our extension of Mangat's RRT approach.

**Keywords** Randomized-response technique · Multinomial model · Social desirability · Dishonest responding · Lie detection

The validity of surveys on sensitive issues suffers from respondents who fail to report their true status with regard to sensitive attributes by responding in line with social norms rather than truthfully (Tourangeau & Yan, 2007). The randomized-response technique (RRT; Warner, 1965) was developed as a means to reduce this social desirability bias. The rationale of the RRT is to add random noise to the participant's responses to ensure that an individual's true status cannot be determined on the basis of his or her response to a sensitive question. Although the randomization process protects the true status of all individuals participating in the survey, elementary probability calculations allow the derivation of estimates of the prevalence of a sensitive attribute at group level.

In the first randomized-response model, proposed by Warner (1965), respondents were to answer either a sensitive question (e.g., "Have you ever used cocaine?") with probability  $p$  or the negation of this question ("Have you never used cocaine?") with probability  $1 - p$ . One implementation of this method simultaneously provided respondents with both the sensitive question and its negation. Participants engaged in a randomization procedure to determine which of the two questions had to be answered. For example, if a die were used as a randomization device, participants could be requested to respond to the sensitive question if the die showed 1, 2, 3, or 4 ( $p = 2/3$ ), but to the negation of the sensitive question if the die showed 5 or 6 ( $1 - p = 1/3$ ). As a consequence, an observed "yes" response might stem from both carriers of the attribute responding to the sensitive question and noncarriers responding to the negation of the sensitive question. The lack

---

M. Moshagen (✉) · J. Musch  
Institute for Experimental Psychology,  
Heinrich Heine University Düsseldorf,  
40225 Düsseldorf, Germany  
e-mail: morten.moshagen@uni-duesseldorf.de

J. Musch  
e-mail: jochen.musch@uni-duesseldorf.de

M. Moshagen · E. Erdfelder  
Psychology III, University of Mannheim,  
68131 Mannheim, Germany

E. Erdfelder  
e-mail: erdfelder@psychologie.uni-mannheim.de

of a direct link between the participant's response and his or her true status regarding the sensitive attributes increases the confidentiality of responses. Given that the probability distribution of the randomization device is known beforehand, it is straightforward to determine the prevalence of the attribute in question. If  $\pi$  denotes the true proportion of respondents carrying the sensitive attribute, the probability of "yes" responses for the entire population is  $\lambda = \pi p + (1 - \pi)(1 - p)$ . It follows that the proportion of respondents carrying the sensitive attribute can be estimated by  $\hat{\pi} = \left[ \hat{\lambda} + (p - 1) \right] / (2p - 1)$ , where  $\hat{\lambda}$  denotes the relative frequency of "yes" responses in a sample and  $p \neq .5$  (Warner, 1965).

Since the randomization procedure adds noise to the responses, the RRT is associated with a considerable loss of efficiency as compared to direct questioning (Lensvelt-Mulders, Hox, & van der Heijden, 2005a). Accordingly, considerable efforts have been exerted in order to reducing the sampling variance (Boruch, 1971; Bourke, 1984; Greenberg, Abul-Ela, Simmons, & Horvitz, 1969; Kuk, 1990; Moors, 1971). One way to achieve this goal is to restrict sampling variability to some part of the total sample. For instance, only one question is asked in the forced-response variant of the RRT (Dawes & Moore, 1980; Greenberg et al., 1969). Rather than requiring a second question, a randomization device is used to determine whether participants are prompted to respond truthfully or instructed to respond in a prespecified way (e.g., "yes"), regardless of the content of the sensitive question. Also addressing the problem of efficiency, Mangat (1994) proposed a variant of the Warner (1965) model that restricts the randomization procedure to participants not carrying the sensitive attribute. In this variant, respondents carrying the sensitive attribute are requested to respond to the sensitive question, whereas respondents not carrying the sensitive attribute receive either the sensitive or the negation of the sensitive question in the format proposed by Warner, as explained above. Given that "yes" responses may still stem from either carriers or noncarriers of the sensitive attribute, the confidentiality of respondents answering affirmatively is protected. Moreover, because only noncarriers are affected by the randomization instruction and all carriers provide "yes" responses, the variance of the estimate of  $\hat{\pi}$  is reduced as compared to Warner's method.

Lack of efficiency, however, is only one of the concerns associated with use of the RRT. A more serious problem pertains to its validity—that is, to the utility of the RRT in obtaining more valid prevalence estimates than a direct question. In particular, the RRT has been criticized as being susceptible to respondents who fail to adhere to the randomized-response instructions by refusing to reply as directed by the randomization device (Campbell, 1987). In every questioning format, including direct questioning,

participants may fail to adhere to the instructions by denying being a carrier of a sensitive attribute in spite of actually being one. In the forced-response variant of the RRT, it is also possible that participants want to avoid associating themselves with an attribute they do not hold, in spite of being told otherwise by the randomization device (Edgell, Duchan, & Himmelfarb, 1992; Lensvelt-Mulders & Boeije, 2007). In fact, a recent meta-analysis has shown that while outperforming direct-questioning formats, the RRT may still substantially underestimate actual population values (Lensvelt-Mulders, Hox, van der Heijden & Maas, 2005b). Thus, although randomization procedures seem to promote more honest responding as compared to direct-questioning formats, the RRT does not completely eliminate the problem of dishonest responding.

Several extensions of the forced-response variant of the RRT have been proposed that allow for determining the proportion of respondents who fail to adhere to the instructions by always providing a self-protecting response (Böckenholt, Barlas, & van der Heijden, 2009; Böckenholt & van der Heijden, 2007; Clark & Desharnais, 1998; Cruyff, van den Hout, & van der Heijden, 2008; Ostapczuk, Moshagen, Zhao & Musch, 2009a; van den Hout, Böckenholt, & van der Heijden, 2010). However, all of these approaches share the limitation that the true status of nonadherent respondents cannot be determined. In other words, these methods estimate the proportion of carriers of the sensitive attribute *not* for the total population of interest, but only for the latent subpopulation of respondents who adhere to the instructions. Unfortunately, the rate of nonadherence may be often substantial, with estimates ranging from 5% to 50% (e.g., Moshagen, Hilbig, & Musch, 2011; Moshagen, Musch, Ostapczuk, & Zhao, 2010; Ostapczuk, Musch, & Moshagen, 2009b; *in press*).

Although the problem of nonadherence to the randomized-response instructions is well recognized and recent developments of the RRT have progressed to the point where the rate of nonadherence can be estimated, there is currently no model available that allows for determining the true status of nonadherent respondents. Existing models either refrain from any assumption regarding the status of nonadherent respondents—with the undesired consequence that researchers cannot draw any conclusion with respect to an often substantial proportion of respondents—or have to rely on questionable assumptions regarding the status of nonadherent respondents that likely result in biased estimates of the prevalence of the sensitive attribute. Unlike existing approaches, the model we propose provides a direct means to estimate the extent of truthful responding, rather than merely estimating the extent of nonadherence to the instructions. Consequently, the proposed model provides undistorted prevalence estimates for

sensitive attributes in the total population of interest, even if respondents fail to respond truthfully.

**A stochastic lie detector**

Mangat’s (1994) variant of the Warner (1965) model restricts the randomization procedure to noncarriers of the sensitive attribute. Noncarriers are prompted to respond to either the sensitive question or its negation, depending on the outcome of the randomization process. In contrast, carriers of the sensitive attribute do not participate in the randomization process and are always requested to respond to the sensitive question. In Mangat’s model, carriers of the sensitive attribute ( $\pi$ ) are assumed to reply truthfully by responding “yes” to the sensitive question with probability 1. Our basic idea is straightforward: We assume that carriers of the sensitive attribute respond truthfully with an unknown probability  $t \leq 1$ , but fail to respond truthfully by responding “no” with probability  $1 - t$ . In line with Mangat’s original model, participants who do not carry the sensitive attribute ( $1 - \pi$ ) are assumed to reply “no” if they receive the sensitive question with probability  $p$  and to reply “yes” if they receive the negation of the sensitive question with probability  $1 - p$ . Figure 1 illustrates our stochastic lie detector model as a tree diagram.

It should be noted that Mangat (1994) also considered the possibility of incomplete truthful responding (i.e.,  $t \leq 1$ ). However, he did so to derive the bias of the  $\hat{\pi}$  estimate obtained through his model under violations of the assumption of complete truthful responding. The difference between our model and Mangat’s original variant is that we use the additional parameter  $t$  to model actual response behavior and, more importantly, to derive undistorted estimates of  $\pi$  even if incomplete truthful responding occurs. Of course, if all carriers respond truthfully ( $t = 1$ ), the stochastic lie detector reduces to Mangat’s (1994)

model, which is therefore a special case of the present approach. Assuming that some participants carrying the sensitive attribute fail to respond truthfully with probability  $1 - t$ , however, the proportion of “yes” responses given a randomization probability  $p_i$  becomes

$$\lambda_i = \pi t + (1 - \pi)(1 - p_i). \tag{1}$$

To estimate the two unknown parameters of the model ( $\pi$  and  $t$ ), it is necessary to draw two independent random samples from the population of interest and to apply different randomization probabilities in each sample ( $p_1 \neq p_2$ ) (Clark & Desharnais, 1998). Specifically, carriers in the first random sample receive the sensitive question with probability  $p_1$  and the negation of the sensitive question with probability  $1 - p_1$ , whereas carriers in the second random sample receive the sensitive question with probability  $p_2$  and the negation of the sensitive question with probability  $1 - p_2$ . To obtain an identifiable model,  $p_1$  and  $p_2$  must differ from each other but need not necessarily sum to 1. By inserting relative frequencies of observed “yes” responses in Eq. 1 separately for the two samples  $i = 1$  and  $i = 2$  ( $\hat{\lambda}_1$  and  $\hat{\lambda}_2$ , respectively) and solving for the parameters, estimates of the parameters can then be derived as:

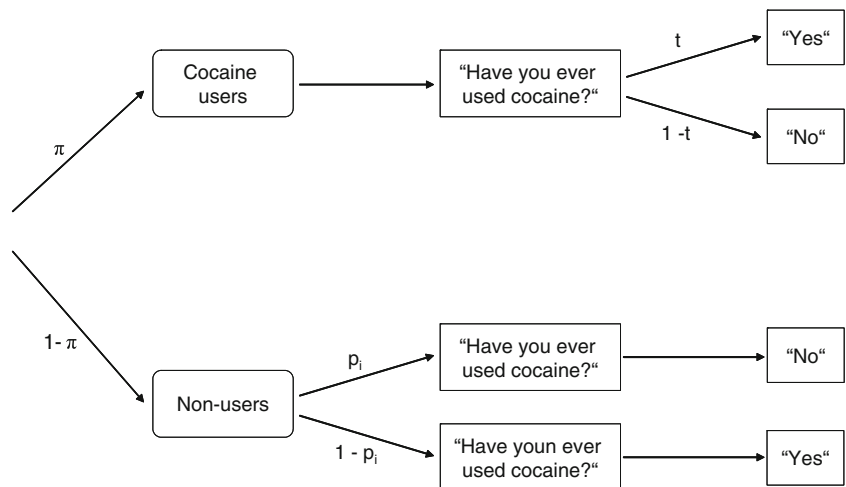
$$\hat{\pi} = \frac{(\hat{\lambda}_2 - \hat{\lambda}_1) + (p_2 - p_1)}{(p_2 - p_1)} \tag{2}$$

and

$$\hat{t} = \frac{\hat{\lambda}_2(1 - p_1) - \hat{\lambda}_1(1 - p_2)}{(\hat{\lambda}_2 - \hat{\lambda}_1) + (p_2 - p_1)}. \tag{3}$$

Equations 2 and 3 provide maximum likelihood estimates of  $\pi$  and  $t$ , respectively, if and only if both estimates lie within the interval  $[0, 1]$ . If either of these estimates lies outside of the admissible interval, this indicates some degree of misfit of the model. The multinomial processing

**Fig. 1** A multinomial processing tree diagram of the stochastic lie detector. Two independent random samples with different randomization probabilities  $p_1$  and  $p_2$  are required to render the model identifiable



tree method outlined below should then be used to estimate parameters and to evaluate the possible significance of the model misfit. Of course, estimates should not be interpreted in case of significant model misfit.

Based on results of Elandt-Johnson (1971, p. 334, Eq. 12.55), the variances of the parameter estimates can be derived as

$$\text{var}(\hat{\pi}) = \frac{\hat{\lambda}_1(1-\hat{\lambda}_1)}{n_1(p_1-p_2)^2} + \frac{\hat{\lambda}_2(1-\hat{\lambda}_2)}{n_2(p_2-p_1)^2} \quad (4)$$

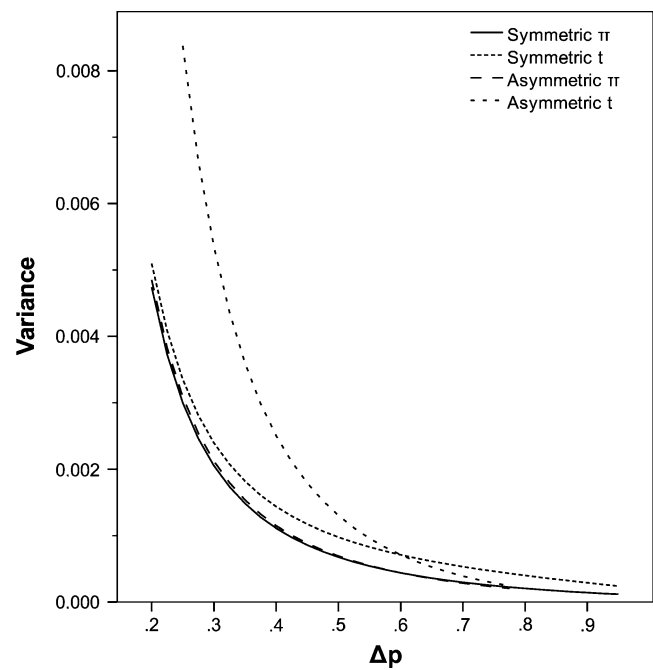
and

$$\text{var}(\hat{t}) = \frac{\hat{\lambda}_1(1-\hat{\lambda}_1)}{n_1} \left[ \frac{(p_1-p_2)(p_2+\hat{\lambda}_2-1)}{(p_1-p_2+\hat{\lambda}_1-\hat{\lambda}_2)^2} \right]^2 + \frac{\hat{\lambda}_2(1-\hat{\lambda}_2)}{n_2} \left[ \frac{(p_1-p_2)(p_1+\hat{\lambda}_1-1)}{(p_1-p_2+\hat{\lambda}_1-\hat{\lambda}_2)^2} \right]^2, \quad (5)$$

where  $n_1$  and  $n_2$  denote the numbers of individuals in the first and second samples, respectively. It is evident from Eq. 4 that the sampling variability of  $\hat{\pi}$  decreases when the denominator terms increase. Because the denominators attain their maximum for a given  $n_i$  when  $|p_1 - p_2| = 1$ , it follows that the sampling variability of the estimates can effectively be reduced by choosing randomization probabilities  $p_1$  and  $p_2$  that differ substantially between the two samples. From a psychological perspective, extreme randomization probabilities should be avoided, because they imply that one sample will almost always have to respond to the sensitive question, counteracting the goal of the RRT to remove any direct link between the true status of a person and his or her response (see Kwan, So, & Tam, 2010, for similar arguments related to a different RRT model)

To further investigate the optimal choice of the randomization probabilities, we conducted a simulation study based on population values of  $\pi = .3$  and  $t = .8$ . The effects of the randomization probabilities were examined in two different ways. In the symmetric case,  $p_1$  and  $p_2$  were varied symmetrically around .5 (such that  $p_1 + p_2 = 1$ ). In the asymmetric case,  $p_1$  took values from .05 to .7, while  $p_2$  was always fixed at .8. From Fig. 2, it can be seen that, as predicted, the variances of  $\pi$  and  $t$  decreased with increasing differences between  $p_1$  and  $p_2$ . Moreover, the variances tended to be slightly smaller with symmetric randomization probabilities, unless the difference between  $p_1$  and  $p_2$  was large.

The closed-form method of parameter estimation based on Eqs. 2 and 3 is just one of several methods to estimate parameters of the stochastic lie detector. Alternatively, we can make use of the fact that the stochastic lie detector is a



**Fig. 2** Variances of  $\hat{\pi}$  and  $\hat{t}$  as a function of the difference between the randomization probabilities  $p_1$  and  $p_2$ , where  $\Delta p = |p_1 - p_2|$ . In the symmetric case,  $p_1$  and  $p_2$  varied symmetrically around .5 such that  $p_1 + p_2 = 1$ . In the asymmetric case,  $p_2$  was fixed at .8 and  $p_1$  was varied from .05 to .7. Estimates are based on  $n_1 = n_2 = 2,500$ ,  $\pi = .3$ , and  $t = .8$

member of the family of multinomial processing tree (MPT) models (Batchelder & Riefer, 1999; Erdfelder, Hilbig, Auer, Aßfalg, Moshagen, & Nadarevic, 2009). Using the MPT approach to estimate the parameters and their associated variances offers several benefits, the main one being increased flexibility. For instance, it is possible to augment the model with additional parameters representing different subgroups (e.g., males and females) for which parameters can be estimated separately. Moreover, in contrast to the estimation procedure based on Eqs. 2 and 3, the MPT approach always provides parameter estimates within the admissible interval  $[0, 1]$ , along with power divergence goodness-of-fit statistics that measure the degree of possible model misfit. By applying the parametric bootstrap test option to one of these goodness-of-fit statistics, it is possible to evaluate the stochastic lie detector statistically even in those cases in which it is a saturated model (with as many parameters as independent model equations—that is, the two-groups case). Likewise, restrictions on the parameters can be tested, such as the assumption that each carrier responds truthfully ( $t = 1$ ). Estimating the stochastic lie detector within the MPT framework proceeds by employing the EM algorithm (Dempster, Laird, & Rubin, 1977) tailored to binary tree models (Hu & Batchelder, 1994). Fortunately, this algorithm has already been implemented in several freely

available software programs (e.g., Hu & Phillips, 1999; Moshagen, 2010; Stahl & Klauer, 2007). Only the program multiTree (Moshagen, 2010), however, provides the parametric bootstrap test option required to evaluate model misfit when one of the parameter estimates obtained with Eqs. 2 and 3 exceeds the permissible interval [0, 1]. The model input file defining the stochastic lie detector model for use in multiTree is shown in the Appendix.

Readers familiar with maximum likelihood procedures for multinomial data can use more general estimation techniques, such as minimizing the log-likelihood ratio statistic  $G^2$  (Read & Cressie, 1988), using software packages such as MATLAB or R. For the present model,  $G^2$  is

$$G^2(t, \pi) = \sum_{i=1}^I 2 \sum_{j=1}^J n_{ij} \log \left( \frac{n_{ij}}{N_i k_{ij}(t, \pi)} \right) \quad (6)$$

where  $n_{ij}$  and  $k_{ij}$  denote the observed and expected frequencies, respectively, for the  $j$ th response category (“yes” and “no” responses) in sample  $i$ , and  $N_i$  is the total number of observations in the  $i$ th sample. This is a very flexible approach that enables users to implement not only different types of parameter restrictions, but also possible extensions of the model that exceed the MPT framework. For example, covariates can be added to the model equations (see, e.g., Böckenholt et al., 2009; Böckenholt & van der Heijden, 2007; van den Hout et al., 2010).

## Experiment 1

An experiment was conducted to assess the performance of the stochastic lie detector in obtaining an undistorted estimate of the prevalence of domestic violence. To this end, we compared the stochastic lie detector with (a) a simple direct question and (b) the RRT variant proposed by Mangat (1994).

A group of 1,504 students (63% female; mean age = 23.57 years,  $SD = 6.86$ ) of the University of Düsseldorf, Germany, volunteered to participate in the study. Respondents were randomly assigned to one of three conditions: direct question ( $N = 329$ ), stochastic lie detector with low randomization probability ( $N = 535$ ), and stochastic lie detector with high randomization probability ( $N = 640$ ). As outlined above, the latter two groups were necessary in order to estimate  $\pi$  and  $t$  based on the stochastic lie detector. Higher numbers of participants were assigned to both stochastic lie detector conditions to compensate for the loss of efficiency associated with the randomization procedure.

Participants in the direct-questioning condition simply received the sensitive statement (“In my current or any former relationship, I have intentionally hurt my partner physically”) and were instructed to respond truthfully. In the stochastic lie detector conditions, participants were provided with both, the

sensitive statement and its negation. The participants’ month of birth was used as a randomization device, to keep the randomization simple and transparent. In the high-randomization-probability ( $p_2$ ) condition, the instructions read:

Below, you find two mutually exclusive statements, which are labeled as statement A and statement B. You are to respond to only one of these statements:

- If you **never** intentionally hurt your current or any former partner physically, please respond to
  - *Statement A*, if you were born between January and October.
  - *Statement B*, if you were born in November or December.
- If you **have** intentionally hurt you current or any former partner physically, simply respond to *Statement A*, regardless of when you were born.

Following these instructions, both the sensitive statement and its negation were presented. They read:

- Statement A: In my current or any former relationship, I have intentionally hurt my partner physically.
- Statement B: In my current or any former relationship, I have never intentionally hurt my partner physically.

The instructions in the low-probability ( $p_1$ ) condition were identical, except that noncarriers were requested to respond to the sensitive statement (Statement A) if they had been born in November or December, and to the negation of the sensitive statement otherwise. According to official birth statistics provided by the German Federal Agency for Statistics, the randomization probabilities  $p_1$  and  $p_2$  were .158 and .842, respectively. Instructions in the stochastic lie detector conditions additionally explained how the randomization procedure protected the confidentiality of responses.

Table 1 presents a summary of the experimental design along with the observed numbers of “yes” and “no” responses

**Table 1** Observed frequencies of “yes” and “no” responses by condition in Experiment 1

	Direct Questioning	Stochastic Lie Detector	
		Low Probability	High Probability
$N$	329	535	630
“Yes”	31	402	215
“No”	298	133	425
$\lambda$	.094	.751	.336
$p$	1	.158	.842

$\lambda$  = relative frequency of “yes” responses;  $p$  = probability of being prompted to respond to the sensitive question

by condition. The parameters of the stochastic lie detector and their associated variances were obtained by applying Eqs. 2–5. Additionally, we employed the MPT framework for parameter estimation using the multiTree program (Moshagen, 2010), because this easily allows for comparing the prevalence estimates of the stochastic lie detector and direct questioning. This is achieved by constraining these prevalence estimates to equality and testing the significance of the associated increase in model misfit ( $\Delta G^2$ ). The MPT framework also makes it possible to compare the performance of the stochastic lie detector with Mangat's (1994) original variant of the RRT. As outlined above, Mangat's method is a special case of the stochastic lie detector based on the assumption that each carrier is responding honestly ( $t = 1$ ).

Table 2 shows the prevalence estimates obtained via direct questioning, Mangat's variant of the RRT, and the stochastic lie detector. The prevalence estimate obtained from the stochastic lie detector indicated that direct questioning underestimated the prevalence of domestic violence. The two procedures yielded significantly different estimates [9.4% vs. 39.3%,  $\Delta G^2$  ( $\Delta df = 1$ ) = 53.26,  $p < .01$ ]. Furthermore, the stochastic lie detector showed that a substantial proportion of carriers failed to respond honestly, since  $t$  was significantly smaller than 1,  $\Delta G^2$  ( $\Delta df = 1$ ) = 56.96,  $p < .01$ .<sup>1</sup> This result shows that relying on Mangat's original variant of the RRT—which does not take dishonest responding into account—would have been inappropriate. Although the prevalence estimate obtained with Mangat's variant of the RRT is significantly higher than the estimate obtained by direct questioning,  $\Delta G^2$  ( $\Delta df = 1$ ) = 8.40,  $p < .01$ , it still underestimates the prevalence of domestic violence as determined by the stochastic lie detector. Hence, application of the stochastic lie detector results in higher prevalence estimates of sensitive issues as compared to both traditional direct questioning and previous randomized-response variants that do not take dishonest responding into account. This provides a first demonstration of the feasibility and utility of the proposed survey method for obtaining undistorted prevalence estimates of sensitive attributes, attitudes, and behaviors.

However, because the true prevalence of domestic violence in the present sample was unknown, one might question whether the higher estimate obtained with the stochastic lie detector was necessarily a more valid estimate. A more precise

**Table 2** Prevalence estimates (standard errors in parentheses) of domestic violence obtained by direct questioning, Mangat's (1994) variant of the randomized-response technique (Mangat), and the stochastic lie detector

	Direct Questioning	Mangat	Stochastic Lie Detector
$\pi$	9.4% (1.6)	17.1% (2.0)	39.3% (3.9)
$t$	–	–	61.1% (3.5)

Mangat's (1994) variant is a special case of the stochastic lie detector, constrained by the assumption that all carriers of the critical attribute respond honestly ( $t = 1$ )

assessment of the validity of the stochastic lie detector would require knowledge of the true status of each individual with regard to a sensitive attribute. Unfortunately, owing to the very nature of a sensitive question, such data are notoriously difficult to obtain. This is mirrored by the fact that a recent meta-analysis by Lensvelt-Mulders et al. (2005b) identified only 6 “strong” validation studies reporting such data, as compared to 68 studies using the comparative validation approach we applied in Experiment 1. However, to achieve a more conservative evaluation of the stochastic lie detector, a second survey was conducted using voting turnout as a better-accessible criterion.

## Experiment 2

The goal of the second experiment was to validate the stochastic lie detector using nonvoting in the 2009 German federal elections as the sensitive behavior. Voter turnout seemed well suited to serve this purpose for three reasons. First, nonvoting in nation-wide elections is a socially undesirable behavior, the prevalence of which is underestimated when relying on direct questioning (Bernstein, Chadha, & Montjoy, 2001; Karp & Brockington, 2005). Second, prior studies employing existing randomized-response models failed to observe improved estimates of voting turnout (Holbrook & Krosnick, 2010; Locander, Sudman, & Bradburn, 1976), suggesting that the RRT cannot always be expected to yield more valid prevalence estimates if untruthful responding is not taken into account. Finally, the true rate of voter turnout is known for the 2009 German federal elections from official statistics, thus enabling its use as an external criterion. For the stochastic lie detector to be considered valid, it must be shown that its estimate of the rate of nonvoting is closer to the known true rate of nonvoting than estimates derived from other forms of questioning. Moreover, because nonvoting arguably is a less sensitive behavior as compared to domestic violence, it was to be expected that carriers of the sensitive attribute would be more inclined to respond honestly (as measured by the parameter  $t$ ) than in the first experiment involving domestic violence.

<sup>1</sup> Parameters at the boundaries of the parameter space violate a regularity condition (Birch, 1964), which questions the asymptotic  $\chi^2(1)$  distribution of  $G^2$  under the null hypothesis. Thus, a parametric bootstrap using 1,000 bootstrap samples was performed (Efron & Tibshirani, 1993). The obtained distribution of the  $G^2$  statistics serves as an estimate of the exact distribution of  $G^2$  that takes into account the boundary value for  $t$ . The parametric bootstrap confirmed the conclusions drawn from asymptotic theory, as the bootstrap distribution yielded  $p < .01$  for observing  $G^2 = 56.96$ , given  $H_0: t = 1$ .

A total of 417 respondents volunteered to participate in the study. Respondents who were not eligible to cast a ballot in the 2009 German parliament election were excluded from further analyses ( $n = 55$ ). Of the remaining participants, 56% were female. Age ranged from 19 to 78 years, with a mean of 31.3 years ( $SD = 11.2$ ). Participants were randomly assigned to one of three conditions: direct questioning ( $N = 66$ ) or stochastic lie detector with low ( $N = 127$ ) or high ( $N = 169$ ) randomization probability, respectively.

The sensitive statement was worded such that nonvoters (carriers of the sensitive attribute) were required to respond affirmatively. It read: “I did not vote in the last Bundestagswahl held in 2009.” Consequently, the negation of the sensitive statement was: “I voted in the last Bundestagswahl held in 2009.” As in [Experiment 1](#), the month of birth was used as the randomization device. Participants in the low-probability ( $p_1$ ) condition who actually voted in the election (noncarriers) were requested to respond to the sensitive statement if they were born between October and December, but to respond to the negation of the sensitive statement otherwise. In the high-probability ( $p_2$ ) condition, noncarriers born between January and September were requested to respond to the sensitive statement, and noncarriers born in another month were prompted to respond to the negation of the sensitive statement. Birth statistics of the German Federal Agency for Statistics showed that the randomization probabilities  $p_1$  and  $p_2$  approximated .251 and .749, respectively. The instructions to the participants closely mirrored those of [Experiment 1](#).

The observed numbers of “yes” and “no” responses by condition are presented in [Table 3](#). The official voting turnout in the 2009 German Bundestag election was 70.8%. Thus, the percentage of eligible voters who did not cast a ballot was 29.2%. The estimated rates of nonvoting for direct questioning, Mangat’s original variant, and the stochastic lie detector are shown in [Table 4](#). It is evident that the rate of nonvoting is substantially underestimated when using a direct-questioning format. In contrast, according to the stochastic lie detector estimate, 31.4% of the eligible voters failed to cast a ballot.

**Table 3** Observed frequencies of “yes” and “no” responses by condition in [Experiment 2](#)

	Direct Questioning	Stochastic Lie Detector	
		Low Probability	High Probability
$N$	66	127	169
“Yes”	6	102	78
“No”	60	25	91
$\lambda$	.091	.803	.462
$p$	1	.251	.749

$\lambda$  = relative frequency of “yes” responses;  $p$  = probability of being prompted to respond to the sensitive question

**Table 4** Prevalence estimates (standard errors in parentheses) of nonvoting obtained by direct questioning, Mangat’s (1994) variant of the randomized-response technique (Mangat), and the stochastic lie detector

Parameter	Direct Questioning	Mangat	Stochastic Lie Detector
$\pi$	9.1% (3.5)	27.3% (4.8)	31.4% (10.5)
$t$	–	–	92.1% (15.7)

This is significantly higher than the direct-questioning estimate,  $\Delta G^2 (\Delta df = 1) = 8.24, p < .01$ , and very close to the true rate of nonvoting,  $\Delta G^2 (\Delta df = 1) = 0.05, p > .8$ . This suggests that estimates obtained with the stochastic lie detector are not only numerically higher than direct-questioning estimates but also more valid. Moreover, participants who did not cast a ballot were more likely to respond truthfully ( $\hat{t} = .92$ ) as compared to carriers of the sensitive attribute in [Experiment 1](#) ( $\hat{t} = .61$ ). Since nonvoting can be considered a less sensitive attribute as compared to domestic violence, this result provides further evidence for the validity of the  $t$  parameter.

Note, however, that the sample of participants questioned in [Experiment 2](#) does not necessarily resemble a random sample drawn from the population of eligible voters. The actual voting turnout of the present sample may therefore have differed from the official voting turnout in the general population. For this reason, the results should not be interpreted as providing strong evidence for the validity of the stochastic lie detector, in the sense defined above. Nevertheless, the close correspondence between the observed pattern of results and the actual election result provides additional support for the validity of the model.

## Discussion

The present contribution addresses problems of biased responding in surveys involving sensitive issues. While the proposed stochastic lie detector, in line with previous randomized-response models, protects the confidentiality of responses using appropriate randomization, it offers the key advantage of providing undistorted prevalence estimates of sensitive issues even when participants fail to respond truthfully. This is an important improvement over alternative approaches that allow for estimating nonadherence to the randomized-response instructions but are not able to unequivocally determine the true status of nonadhering respondents. An additional advantage of the present approach is that, unlike forced-response variants of the RRT, it does not rely on forcing noncarriers of a sensitive attribute to give a self-incriminating response. Many participants in randomized-response surveys find it trou-

blesome to falsely incriminate themselves without actually being a carrier and, in turn, give nonincriminating answers instead (Edgell et al., 1992; Lensvelt-Mulders & Boeije, 2007). In a similar vein, the RRT variant proposed by Mangat (1994), which served as the starting point of our present approach, requires carriers of the sensitive attribute to respond to a direct question. Privacy protection is thereby provided in an indirect way only, as it completely depends on the behavior of noncarriers. This may seriously affect the willingness of carriers to provide truthful responses (Lensvelt-Mulders et al., 2005a). However, it is important to note that, in contrast to other randomized-response models, increased confidentiality due to the randomization procedure is no more than a side effect of using the stochastic lie detector. In fact, given that untruthful responding is taken into account, the stochastic lie detector does not require honest responding at all. Even in the most extreme scenario, in which every single individual carrying the sensitive attribute fails to respond truthfully, the estimate of the prevalence of the sensitive attribute is still undistorted, provided that the stochastic lie detector model is actually valid.

The validity of the proposed model, however, rests on several assumptions. An assumption shared by other questioning models, including the simple case of a direct question, is that noncarriers do not falsely claim to be carriers of the sensitive attribute. Although there is evidence suggesting that such self-incriminating behavior is usually very rare (cf. Tourangeau & Yan, 2007), applied researchers should take caution in applying the model when the direction of social desirability may be ambiguous, as for example when surveying cannabis use among adolescents (e.g., Percy, McAlister, Higgins, McCrystal, & Thornton, 2005).

Another assumption of the proposed model is that carriers will refrain from using the randomization procedure strategically by participating in the randomization procedure and by responding in the same way as noncarriers. However, such a fairly sophisticated response pattern would require fundamental knowledge of the logic of the underlying model in order to successfully hide one's true status. Importantly, the response pattern resulting from such a strategic distortion would result in an underestimate of the prevalence of the sensitive attribute, so that this estimate could still serve as a valid lower bound.

Finally, comparable to related randomized-response models (Clark & Desharnais, 1998; Cruyff et al., 2008; Ostapczuk et al., 2009a), the stochastic lie detector assumes that response behavior does not depend on features of the randomized-response design. Specifically, it is assumed that the probability to respond honestly ( $t$ ) is independent of the randomization probabilities  $p_1$  and  $p_2$ . The validity of this assumption may be questioned, because the degree of

anonymity that is afforded by the model depends on the probability with which noncarriers are instructed to respond to the negation of the sensitive question ( $1 - p$ ). As this probability increases, the diagnostic value of an individual "yes" response decreases (Ljungqvist, 1993). Consequently, carriers may be more willing to respond truthfully in the condition in which the randomization probability is low as compared to the condition in which the randomization probability is high. However, this concern has already been addressed empirically and was not found to be supported by evidence. Soeken and Macready (1982) reported that their participants' willingness to adhere to the instructions was unaffected by different randomization probabilities. In another study, Soeken and Macready (1985) found prevalence estimates of sensitive behaviors to be less rather than more biased by social desirability with an increasing probability of having to answer the sensitive question truthfully. Additionally, it can be shown that violations of the independence assumption concerning  $p$  and  $t$  result in an underestimation of  $\pi$  and an overestimation of  $t$ . Thus, in the worst case, the stochastic lie detector acts conservatively by always underestimating the prevalence of the critical attribute.

The selection of an appropriate randomization device is a critical issue in each application of a randomized-response model. In the present experiments, the participants' month of birth was used. This particular randomization device was chosen because it is simple, transparent, easy to assess and to understand, and impossible to manipulate. Moreover, the month of birth reduces the risk of obtaining a gambling-like situation arising from using a spinner or a die (Moshagen & Musch, *in press*). However, it is vital that the participants trust the integrity of the randomization process. If there is reason to suspect that participants may fear that their month of birth may be known to the investigators, another randomization device should be employed.

Some limitations of the present approach should be acknowledged. Using a randomization procedure to increase anonymity is unavoidably associated with some side effects. First, the prevalence estimates obtained with the stochastic lie detector have a larger variance as compared to those obtained with a direct question. This decrease in efficiency can only be compensated by increasing the sample size. Although it could be argued that obtaining more valid prevalence estimates with less precision might be preferred over obtaining highly precise but invalid prevalence estimates, the additional cost associated with the use of the stochastic lie detector may only be justified when examining issues of a sufficiently sensitive nature as to be severely threatened by socially desirable responding.

Another side effect of the randomization procedure is that the true status of each single individual remains unknown. This makes it difficult to measure associations between the



sensitive attribute and other variables. Whereas some other randomized-response models are capable of including continuous covariates (Böckenholt et al., 2009; van den Hout et al., 2010), the stochastic lie detector has yet to be extended to allow for such an inclusion. Note, however, that measuring associations between the sensitive attribute and categorical predictors (gender, religion, etc.) is less of a problem because the prevalence  $\pi_k$ ,  $k = 1, \dots, K$ , can be estimated separately for each of  $K$  categories or populations.

Finally, the model in its present form is only applicable to dichotomous responses. Although it is straightforward to extend ordinary randomized-response models to other response formats (such as responses on rating scales), it is less clear how to define nonadherence and how to model untruthful responding in these cases (cf. Böckenholt et al., 2009).

Such remaining problems notwithstanding, the stochastic lie detector offers some genuine advantages over previous randomized-response models and traditional direct-questioning formats, both of which yield distorted prevalence estimates to the extent that dishonest responding occurs. We therefore conclude that the stochastic lie detector is a useful extension of RRT methodology for research involving questions on sensitive issues.

**Author Note** We are grateful to Mirjam Horbach for her help in collecting the data.

## Appendix

**Table 5** Presented below is the multiTree (Moshagen, 2010) input file for estimating the parameters of the stochastic lie detector (SLD), as well as the prevalence obtained through direct questioning (DQ), using the multinomial processing tree framework. Y1, Y2, and YD and N1, N2, and ND are the category labels for the numbers of observed “yes” and “no” responses, respectively, in the first SLD, the second SLD, and the DQ conditions, respectively.  $p_i$  and  $d_q$  denote the prevalence estimates obtained through the SLD and DQ, respectively.  $p_1$  and  $p_2$  are the probabilities of receiving the sensitive question in the low- and high-probability conditions, respectively. The probability of responding truthfully is denoted by  $t$

SLD1	Y1	$\pi_i * t$
SLD1	N1	$\pi_i * (1-t)$
SLD1	N1	$(1-\pi_i) * p_1$
SLD1	Y1	$(1-\pi_i) * (1-p_1)$
SLD2	Y2	$\pi_i * t$
SLD2	N2	$\pi_i * (1-t)$
SLD2	N2	$(1-\pi_i) * p_2$
SLD2	Y2	$(1-\pi_i) * (1-p_2)$
DQ	YD	$d_q$
DQ	ND	$(1-d_q)$

## References

- Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin and Review*, 6, 57–86. doi:10.3758/BF03210812
- Bernstein, R., Chadha, A., & Montjoy, R. (2001). Overreporting voting: Why it happens and why it matters. *Public Opinion Quarterly*, 65, 22–44.
- Birch, J. W. (1964). A new proof of the Pearson–Fisher theorem. *Annals of Mathematical Statistics*, 35, 817–824.
- Böckenholt, U., Barlas, S., & van der Heijden, P. G. M. (2009). Do randomized-response designs eliminate response biases? An empirical study of non-compliance behavior. *Journal of Applied Econometrics*, 24, 377–392.
- Böckenholt, U., & van der Heijden, P. G. M. (2007). Item randomized-response models for measuring noncompliance: Risk–return perceptions, social influences, and self-protective responses. *Psychometrika*, 72, 245–262. doi:10.1007/s11336-005-1495-y
- Boruch, R. (1971). Assuring confidentiality of responses in social research: A note on strategies. *The American Sociologist*, 6, 308–311.
- Bourke, P. D. (1984). Estimation of proportions using symmetric randomized response designs. *Psychological Bulletin*, 96, 166–172.
- Campbell, A. (1987). Randomized response technique. *Science*, 236, 1049.
- Clark, S. J., & Desharnais, R. A. (1998). Honest answers to embarrassing questions: Detecting cheating in the randomized response model. *Psychological Methods*, 3, 160–168.
- Cruyff, M. J. L. F., van den Hout, A., & van der Heijden, P. G. M. (2008). The analysis of randomized response sum score variables. *Journal of the Royal Statistical Society: Series B*, 70, 21–30.
- Dawes, R., & Moore, M. (1980). Die Guttman-Skalierung orthodoxer und randomisierter Reaktionen [Traditional Guttman-scaling and randomized response]. In F. Petermann (Ed.), *Einstellungsmessung, Einstellungsforschung* (pp. 117–133). Göttingen: Hogrefe.
- Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39, 1–38.
- Edgell, S. E., Duchan, K. L., & Himmelfarb, S. (1992). An empirical test of the unrelated question randomized response technique. *Bulletin of the Psychonomic Society*, 30, 153–156.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Elandt-Johnson, R. C. (1971). *Probability models and statistical methods in genetics*. New York: Wiley.
- Erdfelder, E., Hilbig, B. E., Auer, T.-S., Abfal, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie / Journal of Psychology*, 217, 108–124.
- Greenberg, B., Abul-Ela, A., Simmons, W., & Horvitz, D. (1969). Unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association*, 64, 520–539.
- Holbrook, A. L., & Krosnick, J. A. (2010). Measuring voter turnout by using the randomized response technique: Evidence calling into question the method’s validity. *Public Opinion Quarterly*, 74, 328–343.
- Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, 59, 21–47. doi:10.1007/BF02294263
- Hu, X., & Phillips, G. A. (1999). GPT.EXE: A powerful tool for the visualization and analysis of general processing tree models. *Behavior Research Methods, Instruments, & Computers*, 31, 220–234. doi:10.3758/BF03207714

- Karp, J. A., & Brockington, D. (2005). Social desirability and response validity: A comparative analysis of overreporting voter turnout in five countries. *Journal of Politics*, *67*, 825–840.
- Kuk, A. (1990). Asking sensitive questions indirectly. *Biometrika*, *77*, 436–438.
- Kwan, S. S. K., So, M. K. P., & Tam, K. Y. (2010). Applying the randomized response technique to elicit truthful responses to sensitive questions in IS research: The case of software piracy behavior. *Information Systems Research*, *21*, 941–959.
- Lensvelt-Mulders, G. J. L. M., & Boeije, H. R. (2007). Evaluating compliance with a computer assisted randomized response technique: A qualitative study into the origins of lying and cheating. *Computers in Human Behavior*, *23*, 591–608.
- Lensvelt-Mulders, G. J. L. M., Hox, J. J., & van der Heijden, P. G. M. (2005a). How to improve the efficiency of randomised response designs. *Quality and Quantity*, *39*, 253–265. doi:10.1007/s11135-004-0432-3
- Lensvelt-Mulders, G. J. L. M., Hox, J. J., van der Heijden, P. G. M., & Maas, C. J. M. (2005b). Meta-analysis of randomized response research: Thirty-five years of validation. *Sociological Methods & Research*, *33*, 319–348. doi:10.1177/0049124104268664
- Ljungqvist, L. (1993). A unified approach to measures of privacy in randomized response models: A utilitarian perspective. *Journal of the American Statistical Association*, *88*, 97–103.
- Locander, W., Sudman, S., & Bradburn, N. (1976). An investigation of interview method, threat and response distortion. *Journal of the American Statistical Association*, *71*, 269–275.
- Mangat, N. (1994). An improved randomized-response strategy. *Journal of the Royal Statistical Society: Series B*, *56*, 93–95.
- Moors, J. (1971). Optimization of the unrelated question randomized response model. *Journal of the American Statistical Association*, *66*, 627–629.
- Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*, 42–54. doi:10.3758/BRM.42.1.42
- Moshagen, M., Hilbig, B. E., & Musch, J. (2011). Defection in the dark? A randomized-response investigation of cooperativeness in social dilemma games. *European Journal of Social Psychology*, *41*, 638–644. doi:10.1002/ejsp.793
- Moshagen, M., & Musch, J. (in press). Assessing multiple sensitive attributes using an extension of the randomized-response technique. *International Journal of Public Opinion Research*.
- Moshagen, M., Musch, J., Ostapczuk, M., & Zhao, Z. (2010). Reducing socially desirable responses in epidemiologic surveys: An extension of the randomized-response-technique. *Epidemiology*, *21*, 379–382.
- Ostapczuk, M., Moshagen, M., Zhao, Z., & Musch, J. (2009a). Assessing sensitive attributes using the randomized-response-technique: Evidence for the importance of response symmetry. *Journal of Educational and Behavioral Statistics*, *34*, 267–287.
- Ostapczuk, M., Musch, J., & Moshagen, M. (2009b). A randomized-response investigation of the education effect in attitudes towards foreigners. *European Journal of Social Psychology*, *39*, 920–931.
- Ostapczuk, M., Musch, J., & Moshagen, M. (in press). Improving self-report measures of medication non-adherence using a cheating detection extension of the randomized-response-technique. *Statistical Methods in Medical Research*.
- Percy, A., McAlister, S., Higgins, K., McCrystal, P., & Thornton, M. (2005). Response consistency in young adolescents' drug use self-reports: A recanting rate analysis. *Addiction*, *100*, 189–196.
- Read, T. R. C., & Cressie, N. A. C. (1988). *Goodness-of-fit statistics for discrete multivariate data*. New York: Springer.
- Soeken, K. L., & Macready, G. B. (1982). Respondents' perceived protection when using randomized response. *Psychological Bulletin*, *92*, 487–489.
- Soeken, K. L., & Macready, G. B. (1985). Randomized response parameters as factors in frequency estimates. *Educational and Psychological Measurement*, *45*, 89.
- Stahl, C., & Klauer, K. C. (2007). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behavior Research Methods*, *39*, 267–273. doi:10.3758/BF03193157
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*, 859–883.
- van den Hout, A., Böckenholt, U., & van der Heijden, P. G. M. (2010). Estimating the prevalence of sensitive behaviour and cheating with a dual design for direct questioning and randomized response. *Journal of the Royal Statistical Society: Series C*, *59*, 723–736.
- Warner, S. (1965). Randomized-response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, *60*, 63–69.