PSYCHOMETRIKA—VOL. 79, NO. 3, 515–537
JULY 2014
DOI: 10.1007/S11336-013-9390-9

MODELING MOTIVATED MISREPORTS TO SENSITIVE SURVEY QUESTIONS

ULF BÖCKENHOLT

NORTHWESTERN UNIVERSITY

Asking sensitive or personal questions in surveys or experimental studies can both lower response rates and increase item non-response and misreports. Although non-response is easily diagnosed, misreports are not. However, misreports cannot be ignored because they give rise to systematic bias. The purpose of this paper is to present a modeling approach that identifies misreports and corrects for them. Misreports are conceptualized as a motivated process under which respondents edit their answers before they report them. For example, systematic bias introduced by overreports of socially desirable behaviors or underreports of less socially desirable ones can be modeled, leading to more-valid inferences. The proposed approach is applied to a large-scale experimental study and shows that respondents who feel powerful tend to overclaim their knowledge.

Key words: response set, survey research, socially desirable responding, self-deceptive enhancement, item response models.

1. Introduction

Many public policy decisions rely on the information extracted from self-reports obtained in surveys or experimental studies. For example, self-reports are indispensable for policy making, developing communication programs and understanding consumer behavior. For the effective use of self-reports, respondents are expected to accurately express their thoughts and beliefs, choices and intentions, as well as many other aspects of themselves and their behaviors. Although self-reports have provided important information in many cases, the limitations of this approach become apparent when studying intentions and private behaviors (Öhman, 2011; Tellis & Chandrasekaran, 2010; Tourangeau & Yan, 2007; van Soest & Hurd, 2008; Wirtz & Kum, 2004). Researchers focusing on such topics as digital-product piracy or consumer cheating on service guarantees as examples of non-compliant behaviors or planned adoptions of new products and donation intentions as examples of future idealistic behaviors cannot rely on the assumption that self-reports are unbiased, candid, and accurate.

In understanding violations of the accuracy assumption in self-reports, the classical distinction of Rorer (1965) between response sets and styles is still of much relevance. Response styles refer to enduring tendencies in answering items that are not specific to the item content. Examples include tendencies to agree with items, to give extreme as opposed to moderate responses, and to give middle or neutral responses (Johnson & Bolt, 2010). In contrast to response styles, response sets are related directly to the item content and refer to the motivation of respondents to answer items in such a way that it facilitates their self-presentation. For example, when measuring compliance with rules and regulations in downloading movies, sharing computer programs, using prescribed products, participating in online dating, or in accurately completing a questionnaire, respondents have been shown to differ not only in their behaviors in these different domains but also in their motivation to present themselves in a positive way (Bowman, Heilman, & Seetharaman, 2004; Harvey & McCrohan, 1988; Mazar & Ariely, 2006; Reingen, 1978; Sinha & Mandel, 2008; Toma, Hancock, & Ellison, 2008; Wosinska, 2005). In the extreme, a person

Requests for reprints should be sent to Ulf Böckenholt, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, USA. E-mail: u-bockenholt@northwestern.kellogg.edu

may be committed to displaying an ideal public self and conveying an image of being flawless (Hewitt, Flett, Sherry, Habke, Parkin, Lam, McMurtry, Ediger, Fairlie, & Stein, 2003). This desire for positive self-presentation may motivate respondents to edit their self-report about the domain under investigation. It also seems likely that the actual behavior can affect the decision on whether to report it without editing. For example, respondents who have been non-compliant may also be highly motivated to hide their behavior. As a result, self-reports about personal and sensitive topics may provide only a biased representation of the actual behaviors.

In this paper, we present a modeling framework to capture response-set effects. The approach is based on the assumption that respondents answer personal questions following a stage-wise process (Tourangeau, Rips, & Rasinski, 2000). First, they arrive at an initial response based on retrieval processes and, subsequently, they decide whether to edit this response and to report a more positive or less revealing answer instead. Several experimental studies on socially desirable responding provide direct support for this conceptualization (Holtgraves, 2004; Wlaczyk, Schwartz, Clifton, Adams, Wei, & Zha, 2005) by showing that social desirability affects the editing process during which participants evaluate retrieved information before responding. The proposed model also assumes that the response-formation and decision-to-edit stages may be correlated in sense that, for instance, individuals whose retrieved response indicates non-compliant behaviors may also be more likely to edit their self-report. Moreover, response-set effects are allowed to be question and person specific because questions may differ in the degree to which they elicit editing, and participants may differ in the degree to which they edit their self-reports. As a result, it is possible to relate covariates to the hypothesized stages, which may prove critical for understanding possible determinants of the response-set effects.

This work contributes to the literature in three ways. First, it operationalizes a motivational framework for modeling response-set effects by introducing editing-decision and response-selection stages subsequently to the formation of the initial response. We incorporate these stages in an item-response model for rating scales and estimate the degree to which an item may be subject to an editing process. Thus, items may differ in the extent to which they elicit editing, and respondents may differ in their motivation to edit their responses. By allowing the editing step to be both item and person specific, the proposed approach goes beyond latent-class models that distinguish between respondents who always edit their responses and those who never do (Böckenholt & van der Heijden, 2007). Second, by controlling for the effects of editing and response selection, we obtain more-valid measurements of the behavior in question. If editing is estimated to play a negligible role in the response process, the proposed approach yields the same results as the corresponding models without this stage. However, the resulting estimates of the sensitive attribute under study can differ substantially when editing matters. Third, from a modeling perspective, the proposed approach joins a growing literature on measurement approaches that, by explicitly accounting for different types of response biases, go beyond the classic view of random sources of measurement error (Benitez-Silva, Buchinsky, Chan, Cheidvasser, & Rust, 2004; Bound, Brown, & Mathiowetz, 2001; Bradlow & Zaslavsky, 1999; Hsiao, Sun, & Morwitz, 2002; Mittal & Kamakura, 2001; Yang, Zhao, & Dhar, 2010).

The remainder of the paper is structured as follows. The next section presents the response-set model and discusses special cases as well as estimation issues. Results of several simulation studies are summarized. We then apply the proposed model to an online study on overclaiming (Paulhus, Harms, Bruce, & Lysy, 2003) and discuss the empirical results. The paper concludes with a discussion of the main findings and avenues for future research.

2. The Retrieve–Edit–Select Framework

The proposed approach separates three potential sources of individual differences in self-reports: (1) behavior as retrieved from memory, (2) the decision for positive self-presentation,

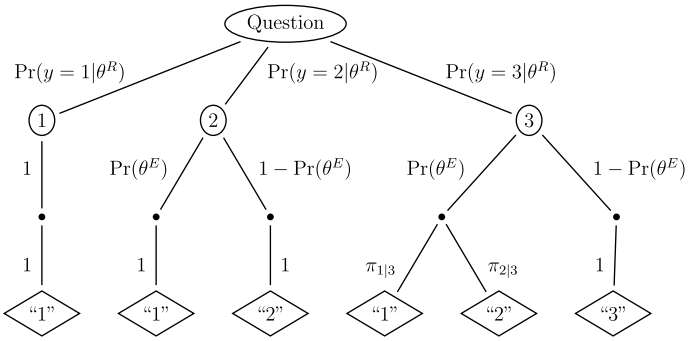

FIGURE 1.
Tree diagram of a sequential response process when a lower response category is less revealing. (Person- and item-specific subscripts are omitted.)

and (3) tendencies in selecting response categories. It is assumed that a person is asked multiple questions about a sensitive or personal domain, each of which may trigger a sequential response process. In the first stage, a person arrives at an initial response retrieved from memory. In the second stage, the respondent decides on whether to edit this response. If this decision is negative, the actual response is reported. However, if the editing decision is positive, the actual response is modified to display the respondent in a more positive way.

Figure 1 illustrates the hypothesized response process for trinary response categories when a lower response is less revealing. The initial responses (denoted by 1, 2, and 3) are listed in circles and the observed ones in rhombuses. For responses 2 and 3, a decision is made whether to edit them. If the decision is positive, a lower response than the initial one is reported. For example, when the initial response is 2 and a person decides to edit it, the observed response is "1". When the initial response is 3, two lower responses can be reported. From Figure 1 is follows immediately that a "3" response can be observed only when a person decides against editing the initial response. We return to this figure after presenting some technical detail about each of the stages.

The outcome of the initial response-retrieval stage is denoted by y_{ij}, which represents the initial response of person i to question j. This response is represented by an item response model for ordinal data with

$$\begin{aligned}\Pr^{(R)}\big(y_{ij} = k|\theta_i^{(R)}\big) &= \Pr{}_{jk}\big(\theta_i^{(R)}\big)\\ &= \Phi\big(\gamma_j^{(R)} + \mathbf{d}_i^{\intercal}\boldsymbol{\varphi}_j^{(R)} + \theta_i^{(R)} - \tau_{k+1}\big)\\ &\quad - \Phi\big(\gamma_j^{(R)} + \mathbf{d}_i^{\intercal}\boldsymbol{\varphi}_j^{(R)} + \theta_i^{(R)} - \tau_k\big),\end{aligned} \tag{1}$$

where R is a short form for the initial response retrieved from memory, $\theta_i^{(R)}$ represents person i's retrieved knowledge or behavior in the sensitive domains under study, $\gamma_j^{(R)}$ is an item effect representing the item location on the sensitive attribute, $\mathbf{d}_i$ is a vector of person-specific covariates with item-specific regression weights $\boldsymbol{\varphi}_j^{(R)}$, and τ_k represents the threshold value of response category k.

The decision outcome of the second stage whether to edit the initial response is denoted by z_{ij} and is represented by a binary probit model with

$$\Pr^{(E)}\big(z_{ij} = 1|\theta_i^{(E)}\big) = \Pr{}_j\big(\theta_i^{(E)}\big) = \Phi\big[\gamma_j^{(E)} + \mathbf{e}_i^{\intercal}\boldsymbol{\varphi}_j^{(E)} + \theta_i^{(E)}\big], \tag{2}$$

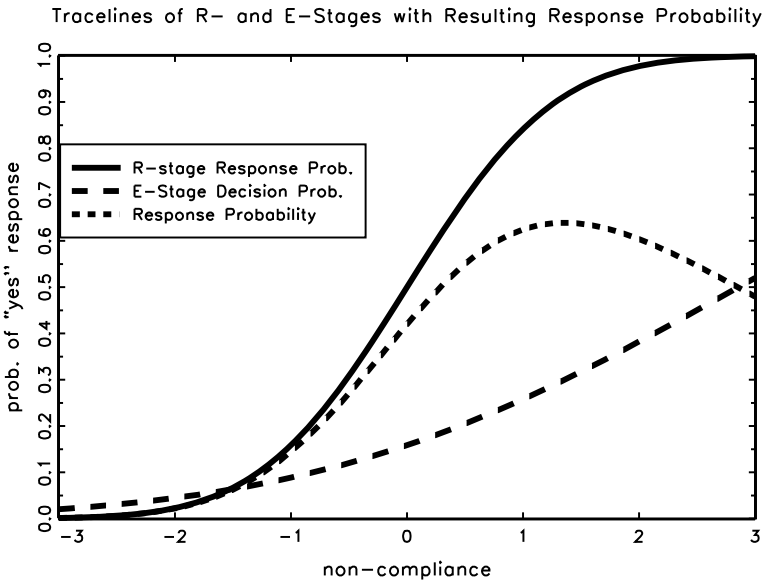

FIGURE 2.
Item tracelines of response stages.

where E stands for the editing decision conditional on the initial response, $\theta_i^{(E)}$ is a person-specific effect representing this person's tendency to edit her or his responses, $\mathbf{e}_i$ is a vector of covariates with corresponding item-specific regression effects $\boldsymbol{\varphi}_j^{(E)}$, and $\gamma_j^{(E)}$ is item j's effect in triggering editing behavior. The conditional nature of the editing decision can be emphasized by writing it as $(E|R)$. However, because there is no confusion about the stage sequence, I refrain from using a more complex superscript.

Finally, the observed self-reports are denoted by x_{ij}. If respondents decide against editing their responses, the observed self-reports x_{ij} are identical to the initial responses y_{ij}. However, if respondents decide to edit their initial responses, it is assumed that they select response categories that are always higher or lower than the corresponding initial ones, depending on the direction that displays a person in a more positive way. These shifts are captured by a transition matrix that relates the initial to the reported response categories. For example, when a rating response lower than the retrieved one is less revealing, the following "editing" matrix is obtained for a five-point rating scale with category labels "1", "2", ..., "5":

$$\boldsymbol{\Pi}\left(\theta_i^{(S)}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \pi_{i1|3} & \pi_{i2|3} & 0 & 0 & 0 \\ \pi_{i1|4} & \pi_{i2|4} & \pi_{i3|4} & 0 & 0 \\ \pi_{i1|5} & \pi_{i2|5} & \pi_{i3|5} & \pi_{i4|5} & 0 \end{pmatrix}. \tag{3}$$

The superscript S denotes the category selection process. The rows of (3) represent the initial rating responses, and the columns, the edited rating responses, with $\pi_{iu|v}$ being the probability that respondent i selects response category u given that the initial response category is v, and $\sum_{k=1}^{K} \pi_{ik|v} = 1$ for $v = 1, \ldots, K$. Note that if the initial rating responses are "1" or "2", the edited responses are always "1".

If a higher response is more desirable, the editing probabilities are complementary to the ones given in (3):

$$\mathbf{\Pi}\left(\theta_i^{(S)}\right) = \begin{pmatrix} 0 & \pi_{i2|1} & \pi_{i3|1} & \pi_{i4|1} & \pi_{i5|1} \\ 0 & 0 & \pi_{i3|2} & \pi_{i4|2} & \pi_{i5|2} \\ 0 & 0 & 0 & \pi_{i4|3} & \pi_{i5|3} \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

Thus, if the initial rating responses are "4" or "5", the edited responses are always "5".

The attractiveness of each response category is estimated by defining $\pi_{ik|v} = \frac{\omega_{ik}}{\sum_{h=1}^{v-1}\omega_{ih}}$ when a lower response is more desirable ($k < v$), and by defining $\pi_{ik|v} = \frac{\omega_{ik}}{\sum_{h=v+1}^{K}\omega_{ih}}$ when a higher response is more desirable ($k > v$). Thus, ω_{ik} captures the attractiveness of the kth response category for respondent i with $\omega_{ik} > 0$. A parsimonious representation of ω_{ik} is obtained by decomposing it into category and person specific parts with $\omega_{ik} = \exp(o_k + \theta_i^{(S)})$, where o_k and $\theta_i^{(S)}$ capture the overall and the person-specific propensities to select category k, respectively. For reasons of identifiability, the first (Kth) response category is set as the reference category when a lower (higher) response is more desirable. For example, $\pi_{i2|4}$ in (3) is specified as

$$\pi_{i2|4} = \frac{\exp(o_2 + \theta_i^{(S)})}{1 + \exp(o_2 + \theta_i^{(S)}) + \exp(o_3 + \theta_i^{(S)})}. \tag{5}$$

To highlight the Retrieve–Edit–Select structure of the proposed model framework, we refer to it as the RES model in the following.

The random effects represented by $\theta_i^{(R)}$, $\theta_i^{(E)}$, and $\theta_i^{(S)}$ are specified to be normally distributed with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$. Dependencies between $\theta_i^{(R)}$ and $\theta_i^{(E)}$ provide insights about whether the editing decision is related or unrelated to the retrieved content. Null correlations between the editing decision and the behavior under study suggest that respondents tend to present themselves in a positive light regardless of their actual behavior. In contrast, non-zero correlations suggest that the initial response and the decision-to-edit have similar underlying causes. For example, non-compliant respondents may also be more likely to edit their responses to hide their behavior.

The branches in Figure 1 list the parameters characterizing the RES stages. A "1" is included to denote that this branch is selected with certainty. The product of the branch probabilities on a path from the "Question" to one of the three observed response categories yields the respective category probabilities. For example, the probability of observing a "3" response by person i to item j can be written as

$$\Pr\left(x_{ij} = \text{"3"}|\theta_i^{(R)}, \theta_i^{(E)}\right) = \Pr_j\left(y_{ij} = 3|\theta_i^{(R)}\right)\Pr_j\left(z_{ij} = 0|\theta_i^{(E)}\right). \tag{6}$$

An illuminating special case is obtained for binary response categories when $\theta_i^{(R)}$ is perfectly correlated with $\theta_i^{(E)}$. Assume that $\theta_i^{(R)}$ represents the actual non-compliance of person i, and $\theta_i^{(E)}$ represents the same person's propensity to edit his responses. Thus, the higher is a person's non-compliance, the higher is her tendency to edit the true response. Figure 2 depicts this scenario with standard deviations 1 and 0.35 for the two random effects $\theta_i^{(R)}$ and $\theta_i^{(E)}$, respectively. Both the response probabilities for the initial response as well as the decision to edit it are monotonic functions of the non-compliance continuum. However, the traceline of the edited-response probabilities is single-peaked. The difference between the initial- and the edited-response probabilities

(and thus observed responses) increases with higher non-compliance levels to the extent that non-compliant respondents become increasingly more likely to give a "no" than a "yes" answer. Clearly, a measurement model under the standard assumption of a monotonic relationship between non-compliance and the probability of a "yes" response would be misspecified and could provide only a biased representation of this process.

2.1. Maximum Likelihood Estimation

Under the scenario that lower responses are less revealing, the conditional distribution of the observed responses given the latent variables $\boldsymbol{\theta}_i = (\theta_i^{(R)}, \theta_i^{(E)}, \theta_i^{(S)})$ can be written as

$$\begin{aligned}
\Pr(x_{ij} = 1|\boldsymbol{\theta}_i) &= \Pr^{(R)}(y_{ij} = 1) + \sum_{k=2}^{K} \Pr^{(R)}(y_{ij} = k)\Pr^{(E)}(z_{ij} = 1)\pi_{i1|k}, \\
\Pr(x_{ij} = k|\boldsymbol{\theta}_i) &= \Pr^{(R)}(y_{ij} = k)\Pr^{(E)}(z_{ij} = 0) + \sum_{c=k+1}^{K} \Pr^{(R)}(y_{ij} = c)\Pr^{(E)}(z_{ij} = 1)\pi_{ic|k}, \quad (7) \\
\Pr(x_{ij} = K|\boldsymbol{\theta}_i) &= \Pr^{(R)}(y_{ij} = K)\Pr^{(E)}(z_{ij} = 0).
\end{aligned}$$

Similar expressions are obtained for the alternative scenario, when higher responses are more desirable. In either case, the log-likelihood function of the observed data $\mathbf{x}$ is given by

$$l(\psi|\mathbf{x}) = \sum_{i=1}^{n}\sum_{j=1}^{J} \int \left[\prod_{k=1}^{K} \left\{\Pr(x_{ij} = k|\boldsymbol{\theta}_i)\right\}^{\delta_{kij}}\right] p(\boldsymbol{\theta}_i|\eta_\theta)\, d\boldsymbol{\theta}_i, \quad (8)$$

where $\delta_{kij} = 1$ if $y_{ij} = k$, and 0 otherwise, for $k = 1, \ldots, K$, and η_θ contains the random-effects distribution parameters. The number of item categories, K, is specified to be at least 3 to allow identifying the fixed and random effects at the S stage. Because for $K = 2$, the E and S stages are confounded, models for binary response categories are not considered here.

Since typically the dimensionality of the latent-variable distribution is three, maximum marginal likelihood methods in combination with Gauss–Hermite quadrature is most effective for the direct maximization of the log-likelihood function given in (8). Specifically, model parameters are estimated by a quasi-Newton method that approximates the inverse Hessian according to the Broyden–Fletcher–Goldfarb–Shanno update (see Gill, Murray, & Wright, 1981). The algorithm utilizes the partial derivatives of the log-likelihood function with respect to all parameters and estimates the Hessian in form of the cross-product of the Jacobian of the gradient. For higher-dimensional models that are obtained when multiple random-effects are specified at each stage, a Monte Carlo EM algorithm is utilized to overcome the numerical problems of numerical quadrature in high-dimensional integrations.

2.1.1. Recovery of Item and Person Parameters Several simulation studies were conducted to assess the performance of the RES model under different specifications and sample sizes. These studies showed that the estimation bias is small. Both the population parameters as well as the corresponding standard errors are recovered reliably at a sample size of $n = 1{,}000$ when the number of items is as small as 5 and the number of response categories is 3. Even for a sample size of $n = 500$, the bias in the parameter estimates is modest. However, the estimated standard errors appear to be systematically smaller compared to the standard deviations of the estimated parameter values. This bias decreases with larger sample sizes. Moreover, RES models with fixed category-attractiveness parameters at the S stage exhibit accurate standard errors when sample sizes are as low as $n = 200$. Details of two simulation studies are reported in Appendix A.

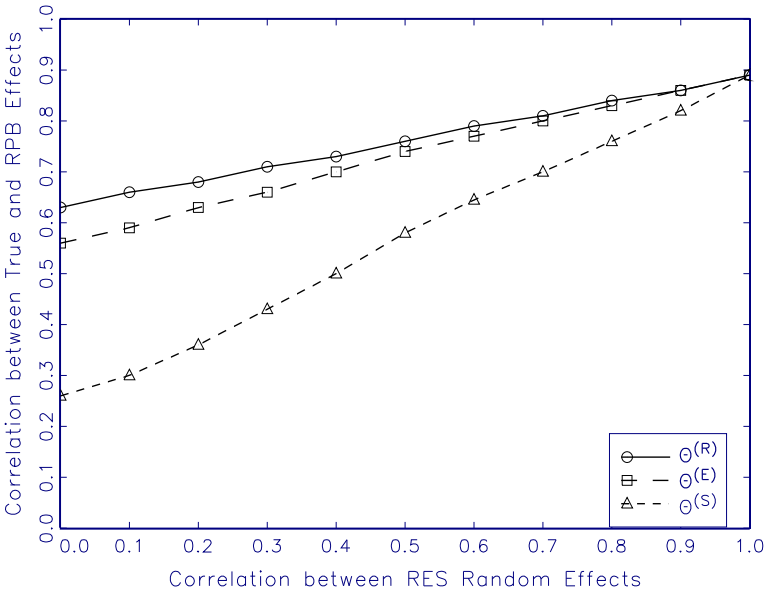

FIGURE 3.
Correlation between true and response-pattern-based (RPB) person effects.

In view of the satisfactory performance in estimating the RES item parameters, we also investigated how well the person-specific measures $\theta^{(R)}$, $\theta^{(E)}$ and $\theta^{(S)}$ of the three response stages can be recovered. This investigation was conducted in two steps. First, we determined the degree to which the discrete nature of the item categories limits a full recovery of the $\boldsymbol{\theta}$ values when the number of items is small. For example, for two items with three response categories each, only the respective means of the θ values for the nine response-category combinations can be estimated. As an upper limit in the recoverability of the θ values, the correlation between the original θ values and the response-pattern-based θ means was determined. In the second step, we assessed the extent to which the RES model can recover the response-pattern-based θ means by computing the correlation between the expected a posteriori (EAP) estimates and the θ means. For this analysis, we focused on the case of five items with three categories each which is also considered in the simulation study reported in Appendix A. This scenario can be viewed as a worst case because more items and more categories per item can only improve the recoverability of the person parameters.

Figure 3 depicts the correlation between the true $\boldsymbol{\theta}_i$ values and the corresponding response-pattern-based $\bar{\boldsymbol{\theta}}_i$ means when the three random effects $\theta^{(R)}$, $\theta^{(E)}$, and $\theta^{(S)}$ are equally correlated. The RES item parameters are identical to the ones chosen for the simulation study reported in Appendix A. Note that the correlation between $\boldsymbol{\theta}_i$ and $\bar{\boldsymbol{\theta}}_i$ increases monotonically and reaches an upper limit of about 0.90 when the random effects are perfectly correlated. When the inter-correlation among the three random effects is low, the correlation between the true $\boldsymbol{\theta}_i$ and the corresponding $\bar{\boldsymbol{\theta}}_i$ is about 0.60 for the R and E stages but only 0.35 for the S-stage. In view of Figure 1, the low correlation for $\theta^{(S)}$ is not surprising. Only respondents who select category "3" and decide to edit their responses provide information about $\theta^{(S)}$.

Table 1 reports the mean correlations among the true $\boldsymbol{\theta}_i$ values, the corresponding $\bar{\boldsymbol{\theta}}_i$ values as well as the EAP estimates from the simulation study reported in Appendix A. For $n = 5{,}000$ the correlation for the three pairs $(\theta^{(R)}, \bar{\theta}^{(R)}) = 0.77$, $(\theta^{(E)}, \bar{\theta}^{(E)}) = 0.75$ and $(\theta^{(S)}, \bar{\theta}^{(S)}) = 0.59$ is also depicted in Figure 3 for a correlation of 0.5 among the random effects. The next two columns in Table 1 present the correlations between the true $\boldsymbol{\theta}_i$ values and the EAP estimates

TABLE 1.
Correlations between true, response-pattern-based (average), and estimated person-specific effects for simulation study in Appendix A.

Stage	$n = 5{,}000$			$n = 1{,}000$			$n = 500$		
	$r(\theta, \bar{\theta})$	$r(\theta, \hat{\theta})$	$r(\bar{\theta}, \hat{\theta})$	$r(\theta, \bar{\theta})$	$r(\theta, \hat{\theta})$	$r(\bar{\theta}, \hat{\theta})$	$r(\theta, \bar{\theta})$	$r(\theta, \hat{\theta})$	$r(\bar{\theta}, \hat{\theta})$
R	0.77	0.68	0.87	0.81	0.68	0.84	0.83	0.68	0.82
E	0.75	0.67	0.89	0.78	0.67	0.85	0.81	0.66	0.82
S	0.59	0.53	0.90	0.65	0.53	0.82	0.70	0.53	0.76

and the correlations between the response-pattern-based $\bar{\boldsymbol{\theta}}_i$ values and the EAP estimates, respectively. The recovery at the response-pattern level is satisfactory with an average correlation of about 0.90. The table also reports the respective correlations for $n = 1{,}000$ and $n = 500$. Since not all response patterns are observed for these two sample sizes, the $\bar{\boldsymbol{\theta}}_i$ values are poorly estimated with the result that the correlations for $(\boldsymbol{\theta}_i, \bar{\boldsymbol{\theta}}_i)$ and for $(\bar{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\theta}}_i)$ are over- and underestimated, respectively. However, for each sample size the relationship $r(\theta, \hat{\theta}) = r(\theta, \bar{\theta}) r(\bar{\theta}, \hat{\theta})$ is satisfied.

Overall, these result show that even for a small number of items and response categories, person-specific effects can be estimated. Modeling individual differences in the category selection probabilities may be most useful when the correlations between $\theta^{(S)}$ and the other two random effects $\theta^{(R)}$ and $\theta^{(E)}$ are substantial or when the number of items and response categories is large. If the intercorrelations are low and the number of items is small, little may be gained by including random effects at this stage.

3. Measuring Self-enhancing via Overclaiming

The following application uses the overclaiming technique (Paulhus et al., 2003) to investigate whether the proposed RES model can provide a parsimonious representation of response bias. The overclaiming technique operationalizes the degree to which respondents may self-enhance by asking them to assess their familiarity with such topics as events or products. Some of the topics are real, but others are made up. Claiming to be familiar with the fictitious items is indicative of self-enhancing behavior. Recent support for the potential of the overclaiming method in applied work was provided by Tellis and Chandrasekaran (2010) who used purchase claims of made-up products as indicators of socially desirable responding. In their study, over 40 % of the respondents claimed to have owned a product that did not exist.

The main purpose of our investigation is to test whether the postulated editing stage of the RES model can identify tendencies of participants to overclaim their knowledge while controlling for their familiarity with the domain under study. To this end, we experimentally induced feelings of power, defined as the asymmetric control over other people or valued resources (Magee & Galinsky, 1992). When people experience power, as opposed to lacking power, they are more likely to rely on their own thoughts (Brinõl, Petty, Valle, Rucker, & Becerra, 2007), and they become more sensitive to opportunities that allow them to display superiority (Campbell, Goodie, & Foster, 2004). Thus, participants who are primed with power may present themselves as being more knowledgeable and competent than they would do otherwise.

In addition to the power manipulation, we also measured socially desirable responding with the self-deceptive enhancement (SDE) scale of the BIDR instrument (Paulhus, 2002). This scale captures tendencies to exaggerate one's social and intellectual status (Paulhus, 2002). If knowledge overclaiming is motivated by goals of exaggerating one's intellectual status, responses to the SDE scale may be correlated with individual differences at the editing stage of the RES model.

As a potential predictor of individual differences in the initial familiarity assessments, the participants' need for cognition (NfC) (Cacioppo & Petty, 1982) was measured. NfC represents a dispositional tendency toward analytical thought and has been shown to be a stable personality variable (Sadowski & Guloz, 1992) that is central to understanding different components of information processing and behavior. For example, people in high need for cognition are more likely to form their attitudes by paying close attention to relevant arguments and are less likely to be affected by contextual and framing effects than people in low need for cognition (Simon, Fagley, & Halleran, 2004; Cacioppo, Petty, Feinstein, & Jarvis, 1996). Thus, participants with high NfC scores may be expected to be more knowledgeable but not to be more prone to socially desirable responding than participants with low NfC scores.

3.1. Method

Respondents were 514 participants in a Canadian online research panel who received monetary compensation for taking part in the study. Participants in the panel were predominantly female (66.2 %) and were distributed across different age groups as follows: 7 % were 20 or younger, 30 % were between ages 21 and 30, 29 % were between 31 and 40, 18 % were between 41 and 50, and the remaining 16 % were over 50. All of the participants had expressed a personal interest in the life sciences as part of a screening question.

The study consisted of two between-subject conditions. In the first ("Power") condition, respondents were asked to recall an event in which they experienced power. In the second condition, hereinafter referred to as "Control", respondents were asked to recall an event in which they went shopping. Power was manipulated via an episodic prime adapted from Galinsky, Gruenfeld, and Magee (2003). Specifically, respondents were instructed to recall an event: "For this task, we are interested in the words people use when reporting on past events as well as the language people generally use to describe past events. On the next screen, you will be asked to recall an event from your life. Please recall the event using the words and language you normally would". In the power condition, the instructions continued as follows: "Please recall a particular incident in which you had power over another individual or individuals. By power, we mean a situation in which you controlled the ability of another person or persons to get something they wanted, or were in a position to evaluate those individuals. Please describe this situation in which you had power—what happened, how you felt, etc." In the control condition, participants read: "Please recall a particular incident in which you went to the grocery store. Please describe this situation in which you went to the grocery store—what happened, how you felt, etc."

Subsequently, all participants self-assessed their familiarity with 12 real and three made-up terms from the life sciences (Paulhus et al., 2003). They also filled out the SDE part of the BIDR questionnaire, the NfC questionnaire (Cacioppo, Petty, & Kao, 1984) as well as a manipulation check and a multiple-choice test on the terms from the overclaiming questionnaire (see Appendix B). The multiple-choice test was included to obtain a more objective measure of the respondents' knowledge level than that provided by the familiarity ratings. The study took approximately 15–20 minutes to complete. As a reward for their participation, all respondents entered a lottery for cash prizes.

3.2. Descriptive Results

Although by using a large number of items we can learn more about individual differences in self-enhancing behavior, typically this strategy is too costly to implement in many studies. For this reason, it is important to show that even for a small number of items, socially desirable responding can be identified and possibly be adjusted for. To this end, I selected three real items (sciatica, meiosis, and antigen) and three made-up items (meta-toxins, bio-sexual and retroplex) from the overclaiming questionnaire. Participants assessed each of the six items on a five-point

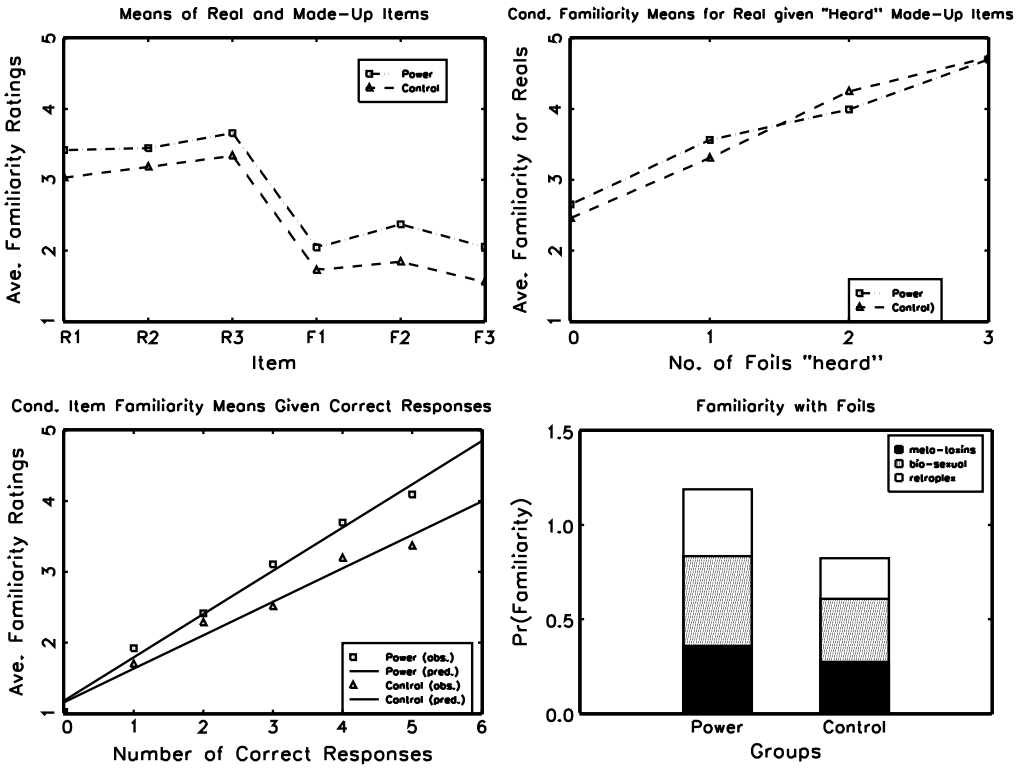

FIGURE 4.
Descriptive results under Power and Control conditions. The *top-left panel* depicts the six items (three reals (R) and three foils (F)) under the Power and Control conditions. The *top-right panel* displays the means of the three real items given the number of made-up items to which participants gave at least a "heard" ratings. The *bottom-left panel* shows the regression lines of the familiarity ratings given the number of correctly answered items. The *bottom-right panel* depicts the relative frequency of respondents claiming to be familiar with the three made-up items in the Power and Control conditions.

scale with the categories (1) "never heard", (2) "heard", (3) "somewhat familiar", (4) "familiar", and (5) "very familiar". The top-left panel of Figure 4 displays the six item means for the two conditions. As expected, participants were more familiar with the real than with the made-up items. The mean differences between the Power and the Control conditions are significant ($F(6, 507) = 4.19$, $p < 0.001$). Moreover, there is no evidence of an item by condition interaction ($F(5, 508) = 0.93$, $p > 0.4$).

The top-right panel of Figure 4 shows that responses to the made-up items are predictive of the responses to the real items. This figure displays the average ratings of the three real items, taking into account the number of made-up items to which a participant gave at least a "heard" response. For both conditions, the regression function of the familiarity responses with respect to the real items given the number of "heard of" non-existent items is almost perfectly linear. This result suggests that individual differences in the familiarity ratings are similar for the real and made-up items.

Over 40 % of the participants claimed to have heard of or to be familiar with the foils. The bottom-right panel of Figure 4 depicts the relative frequencies for the three foils in both conditions. It is apparent that more respondents claim familiarity with the made-up items in the Power than in the Control condition ($\chi^2 = 24.3$, $\mathrm{df} = 7$, $p < 0.001$).

The bottom-left panel of Figure 4 illustrates how knowledge about the domain may moderate the overclaiming effect. The familiarity means are plotted conditional on the number of

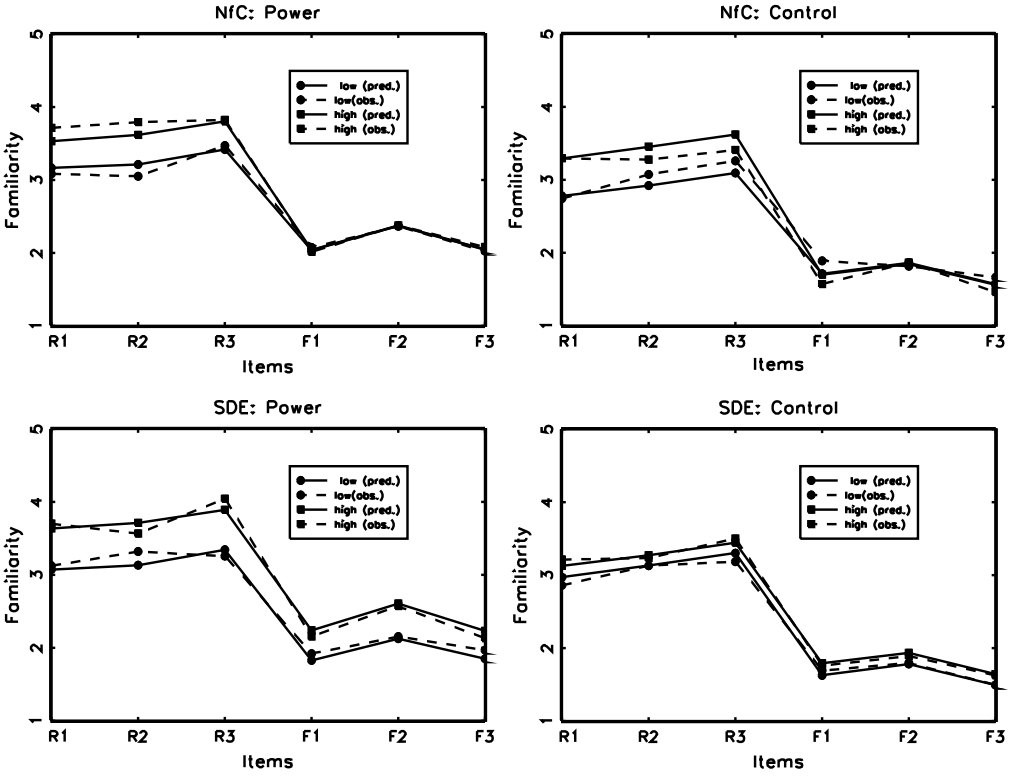

FIGURE 5.
Model-based predictions for Power and Control conditions. The four panels depict the six item means (three reals (R) and three foils (F)) for median-split values of the covariates "Need for Cognition" (NfC) and "Self-Deception Enhancement" (SDE) under the Power and Control conditions. The *continuous* and *dashed lines* connect the observed and predicted means, respectively. The *top panels* display the effect of NfC under the Power and Control condition. The *bottom panels* display the corresponding effects of SDE.

correctly answered items in the multiple-choice test. As expected, the more items are answered correctly, the more a person is familiar with the items. However, the linear regression lines fitted to these means differ between conditions. The slope is significantly steeper ($t = 4.84$, $\mathrm{df} = 11$, $p < 0.001$) for familiarity claims in the Power than in the Control condition, showing that more-knowledgeable respondents exhibit an even stronger self-enhancement effect than less-knowledgeable respondents. Although not displayed in this figure, the number of correctly answered items does not appear to differ between conditions ($\chi^2 = 0.4$, $\mathrm{df} = 7$, $p > 0.5$), suggesting that there are no knowledge differences between the two conditions.

The *solid* lines in Figure 5 depict the relationships between the median-split NfC and SDR measures and the familiarity ratings. In the top panels higher NfC scores are associated with higher familiarity ratings for the three real but not for the three made-up items. In contrast, the bottom panels show that higher SDE scores are associated with higher familiarity ratings across both real and made-up items. However, the strength of this association seems to depend on the condition. In the Power condition the SDE effect seems to be stronger than in the Control condition. The dashed lines in Figure 5 are the predicted means under the RES model. The estimation results of this model are presented in the next section.

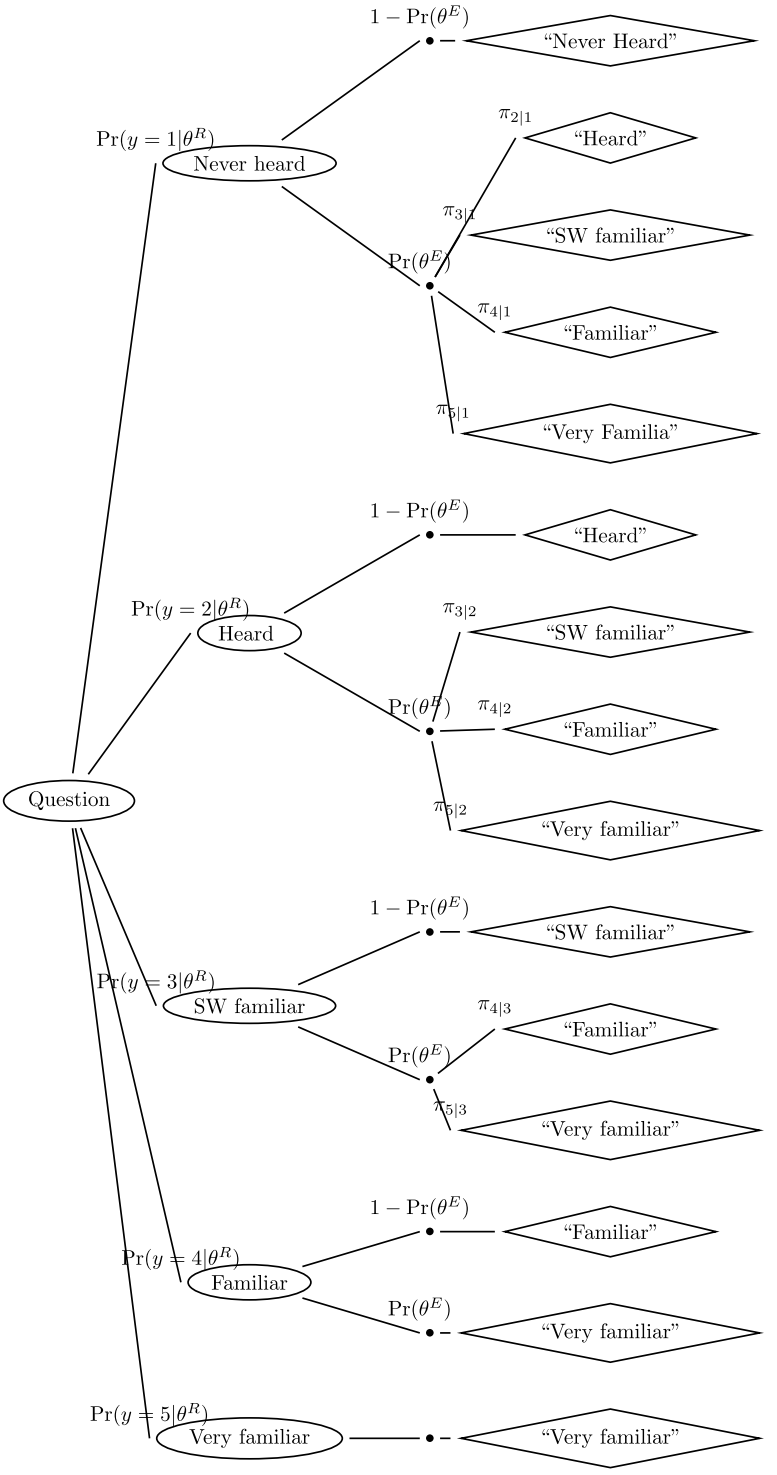

FIGURE 6.
Tree diagram of postulated response process for overclaiming items. *Note*: The notation omits person- and item-specific subscripts. Branches without a parameter are selected with a probability equal to 1.

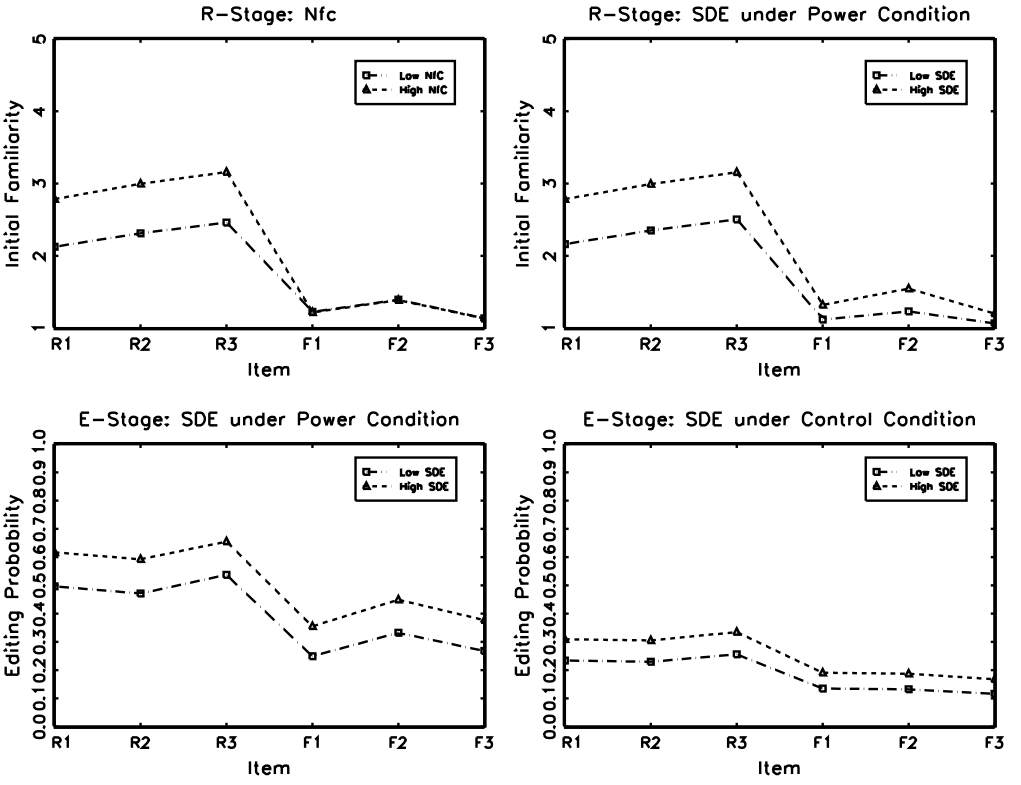

FIGURE 7.
Model-based predictions for "Need-for-Cognition" and "Self-Deception Enhancement" effects. The four panels illustrate the effects of the "Need-for-Cognition" (NfC) and "Self-Deception Enhancement" (SDE) covariates on the R and E stages of the RES model. The *top left* and *right panels* depict the estimated item means at the R stage for median-split NfC and SDE scores, respectively. The *bottom panels* display the estimated average editing probabilities for median-split SDE scores under the Power (*left panel*) and the Control (*right panel*) conditions.

3.3. Estimation Results of the RES Model

Figure 6 depicts the response process postulated under the RES model. Each item is assumed to give rise to an initial familiarity level that is either reported or edited. In the latter case, a higher than the initial rating category is selected. Since each of these stages is assumed to give rise to individual differences, they can be related to the available NfC and SDE measures. In view of the descriptive results obtained previously, it is of interest to determine whether NfC scores predict individual differences in the initial familiarity ratings and whether SDE scores predict the respondents' decision to edit them. Such a finding would be consistent with stage-specific experimental effects in that the R stage parameters are unaffected by the experimental manipulations and the incidence of editing is higher under the Power than the Control conditions.

3.3.1. Model Selection and Tests To test these predictions, the most general RES model, fitted to each condition separately, was specified to have different item parameters and random effects for each of the three stages. It also included stage-specific regression effects with $\varphi_1^{(R(r))}\mathrm{NfC}_i + \varphi_2^{(R(f))}\mathrm{NfC}_i + \varphi_3^{(R)}\mathrm{SDE}_i$ at the R stage and $\varphi_1^{(E)}\mathrm{NfC}_i + \varphi_2^{(E)}\mathrm{SDE}_i$ at the E stage. The superscripts in $\varphi_1^{(R(r))}$ and $\varphi_2^{(R(f))}$ indicate that the regression effect depends on whether the item is real (r) or a foil (f). This RES model yielded a log-likelihood of $-3{,}589.7$ with 58 parameters (see Table 2 for a list of the parameters).

TABLE 2.
Estimates of RES model.

Conditions:	Power			Equal			Control		
Effects	Est.	SE	Est/SE	Est.	SE	Est/SE	Est.	SE	Est/SE
$\gamma^{(R)}_{\text{sciatica}}$				0.25	0.14	1.84			
$\gamma^{(R)}_{\text{meiosis}}$				0.45	0.14	3.20			
$\gamma^{(R)}_{\text{antigen}}$				0.61	0.17	3.62			
$\gamma^{(R)}_{\text{meta-toxins}}$				−1.64	0.23	−7.00			
$\gamma^{(R)}_{\text{bio-sexual}}$				−1.18	0.21	−5.59			
$\gamma^{(R)}_{\text{retroplex}}$				−1.99	0.28	−7.01			
$\gamma^{(E)}_{\text{sciatica}}$	0.24	0.22	1.11				−1.01	0.36	−2.81
$\gamma^{(E)}_{\text{meiosis}}$	0.14	0.25	0.54				−1.04	0.39	−2.66
$\gamma^{(E)}_{\text{antigen}}$	0.41	0.28	1.46				−0.90	0.49	−1.85
$\gamma^{(E)}_{\text{meta-toxins}}$	−0.86	0.18	−4.87				−1.64	0.30	−5.49
$\gamma^{(E)}_{\text{bio-sexual}}$	−0.46	0.19	−2.42				−1.66	0.33	−5.04
$\gamma^{(E)}_{\text{retroplex}}$	−0.76	0.16	−4.78				−1.79	0.32	−5.54
τ^*_1				−0.68	0.10	−6.50			
τ^*_2				−0.45	0.12	−3.94			
τ^*_3				−1.22	0.36	−3.36			
o^*_1				−0.58	0.49	−1.19			
o^*_2				−0.43	0.46	−0.94			
o^*_3				0.78	0.36	2.17			
$\varphi^{(R(r))}_{\text{NfC}}$				0.42	0.08	5.01			
$\varphi^{(R(f))}_{\text{NfC}}$				−0.02	0.11	−0.16			
$\varphi^{(E)}_{\text{NfC}}$				0.09	0.12	0.70			
$\varphi^{(R)}_{\text{SDE}}$	0.43	0.10	4.31				0.01	0.10	0.09
$\varphi^{(E)}_{\text{SDE}}$	0.31	0.11	2.85				0.25	0.13	1.92

Note: $\tau_k = \sum_{f=2}^{k} \exp(\tau^*_f)$ and $\tau_1 = 0$. Parameter estimates that are equal between conditions are listed under the "Equal" column. The covariance-matrix estimates are presented in the text.

Subsequently, special cases of this general model were estimated to test for differences between the conditions. We found that the item parameters of the R stage, the category-selection parameters of the S stage, and the item threshold parameters did not vary across conditions. However, the item parameters of the E stage differed systematically between the Power and Control conditions. Moreover, the NfC scores predicted variations in the R stage but not in the E stage, whereas the SDE scores predicted variations in both the E and the R stages. The log-likelihood of the simplified RES model is −3,597.4 with 43 parameters (see Table 2). This model fit does not differ significantly from the one of the general RES model ($\chi^2 = 15.3$, df = 15).

To determine the fit improvement provided by the E and S stages, we also estimated an ordinal item-response model (Samejima, 1997) with an R stage only. To make this model comparable to the parameterization of the RES model, different item and slope parameters were estimated for each condition; and the same constraints were imposed on the effects of the SDE and NfC covariates as for the simplified RES model. The parameter estimates of this model are presented in Table 3. The table shows that the item effects are significantly different between conditions.

TABLE 3.
Estimates of ordinal item response model.

Conditions:	Power			Equal			Control		
Effects	Est.	SE	Est/SE	Est.	SE	Est/SE	Est.	SE	Est/SE
γ_{sciatica}	0.93	0.10	9.61				0.60	0.09	6.87
γ_{meiosis}	0.97	0.10	9.92				0.77	0.09	8.72
γ_{antigen}	1.19	0.10	12.03				0.87	0.09	9.81
$\gamma_{\text{meta-toxins}}$	−0.47	0.10	−4.53				−0.68	0.10	−6.82
$\gamma_{\text{bio-sexual}}$	−0.10	0.10	−0.92				−0.52	0.10	−5.47
$\gamma_{\text{retroplex}}$	−0.47	0.10	−4.52		–		−0.92	0.10	−8.75
τ_1^*				−1.12	0.06	−18.32			
τ_2^*				−1.02	0.06	−17.81			
τ_3^*				−0.51	0.05	−11.31			
$\varphi_{\text{NfC}}^{(r)}$				0.25	0.05	5.01			
$\varphi_{\text{NfC}}^{(f)}$				−0.04	0.05	−0.77			
φ_{SDE}	0.32	0.06	4.94				0.09	0.06	1.51

Note: $\tau_k = \sum_{f=2}^{k} \exp(\tau_f^*)$ and $\tau_1 = 0$. Parameter estimates that are equal between conditions are listed under the "Equal" column.

TABLE 4.
Orlando–Thissen χ^2 fit statistics.

Item	Power				Control			
	1(2345)	(12)(345)	(123)(45)	(1234)5	1(2345)	(12)(345)	(123)(45)	(1234)5
sciatica	1.6	0.9	7.4	5.3	3.7	3.7	3.1	12.2
meiosis	1.9	1.4	4.2	1.4	2.1	3.9	4.0	7.2
antigen	1.1	3.2	9.3	1.2	1.5	5.8	6.6	10.4
meta-toxins	1.2	2.3	6.3	6.1	3.1	4.7	9.6	5.4
bio-sexual	2.7	1.8	2.4	2.3	2.7	1.3	1.6	6.1
retroplex	2.3	3.2	3.0	4.1	4.2	1.7	2.1	5.1

Moreover, the SDE and NfC effects are similar to the ones obtained for the *R* stage of the RES model. However, the model fits significantly worse than the RES model with a log-likelihood of −3,664.5 and 21 parameters. It is noteworthy that the ordinal item-response model may be viewed as an extension of the signal-detection model (Swets, 1964) which often is used for the analysis of overclaiming data (Paulhus et al., 2003). The poorer fit of the ordinal regression model suggests that the signal-detection model may not provide a full description of overclaiming responses.

The overall fit of the RES model was assessed by computing the Orlando–Thissen χ^2 fit statistics (Orlando & Thissen, 2000) for each item by grouping successive response categories into pairs (denoted by 1(2345), (12)(345), (123)(45) and (1234)5). Table 4 shows that none of the items appears to exhibit systematic misfit across the two conditions when compared to a χ^2-distribution with four degrees of freedom. Tests of the covariates also indicated a satisfactory fit of the data. This fit is illustrated by the dashed lines in Figure 5, which show that the effects of the covariates are well captured by the RES model.

3.3.2. RES Model Predictions The estimated parameters in Table 2 describe the separate effects of each of the RES stages. Figure 7 visualizes these effects. The top-left panel of this figure depicts the estimated item means at the *R* stage for median-split NfC scores. These values

are lower than the observed ones, indicating a substantial degree of editing. The "never heard" category is the modal response for the made-up items at this stage and is estimated to be selected by 88 % of the respondents. The model also predicts that NfC scores account for individual differences at the R stage for the real but not for the made-up items. The top-right panel shows the corresponding effects for median-split SDE scores under the Power condition. Here, the SDE scores predict respondents experiencing higher familiarity at the initial stage with a proportionally stronger effect for real than for made-up items. Importantly, this effect is found in the Power condition only, suggesting that the power prime increased attention paid to the initial familiarity assessments.

The bottom panels of Figure 7 depict the items' average decision-to-edit probabilities for mean-split SDE scores under the two conditions. These plots show clearly that the power-prime manipulation increased tendencies to edit the initial ratings. The average editing probabilities under the Power and Control conditions are 0.45 and 0.21, respectively, with higher editing probabilities for the real than for the made-up items. This result suggests that, if the goal is to measure editing behavior, real items may be more informative than made-up items because made-up items appear to trigger less editing. The plots also highlight that SDE scores account for individual differences in the editing decision but less so in the Control than in the Power condition. From the similar profiles of the average initial familiarity ratings and of the average probabilities to edit, we conclude that initial ratings of more-familiar items are more prone to be edited than less-familiar ones.

The aggregate estimated category-selection probabilities of the S stage are listed in (9), showing that, on average, "familiar" is picked as the modal response category when the decision to edit is made:

$$\hat{\mathbf{\Pi}}\left(\theta^{(S)}\right) = \begin{pmatrix} 0 & 0.12 & 0.14 & 0.47 & 0.27 \\ 0 & 0 & 0.16 & 0.54 & 0.30 \\ 0 & 0 & 0 & 0.66 & 0.34 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{9}$$

Finally, the covariance matrices (with correlations in the upper diagonal) of $\theta_i^{(R)}$, $\theta_i^{(E)}$, and $\theta_i^{(S)}$ under the Power and Control conditions are estimated as

$$\hat{\mathbf{\Sigma}}^{(\text{Power})} = \begin{pmatrix} 0.77 & 0.65 & 0.56 \\ 0.74 & 1.67 & 0.60 \\ 0.50 & 0.80 & 1.05 \end{pmatrix}, \tag{10}$$

and as

$$\hat{\mathbf{\Sigma}}^{(\text{Control})} = \begin{pmatrix} 1.02 & -0.22 & -0.76 \\ -0.29 & 1.72 & 0.73 \\ -0.58 & 0.72 & 0.57 \end{pmatrix}, \tag{11}$$

respectively. Thus, the individual-difference effects, $\theta_i^{(E)}$ and $\theta_i^{(S)}$, are positively correlated in both conditions (0.60 and 0.73, respectively), suggesting that the decision to edit and the selection of a response category are influenced by similar factors. In contrast, $\theta_i^{(R)}$ is positively correlated with $\theta_i^{(E)}$ in the Power condition only. The latter result indicates that under the power manipulation, respondents who experienced higher familiarity in the R stage were also more prone to edit their ratings. The lower-left panel of Figure 4 provides support for this interpretation, showing that the relationship between the number of correctly answered items and the familiarity rating is stronger for respondents in the Power than in the Control condition.

3.3.3. Model Validation Several results speak to the postulated stage structure of the RES model. First, the item parameters at the E stage vary systematically between the power and control conditions indicating that, on average, participants are more likely to edit their familiarity response when they get primed with power. Moreover, SDE scores predict individual differences at the editing stage with similar effects for the power and the control conditions. Since the SDE scale measures tendencies to exaggerate one's intellectual status, these associations provide validation for the interpretation of this stage. Interestingly, SDE is also correlated with individual differences at the R stage in the power but not in the control condition. This result suggests that respondents primed with power may have assessed their familiarity with more confidence than respondents in the control condition.

In contrast to the SDE scores, NfC scores were mainly correlated with individual differences at the R stage. Respondents who scored higher on the NfC scale expressed more familiarity with the real but not with the fictitious items. This result is consistent with Figure 5, which shows that the NfC scores are not predictive of familiarity differences for the made-up items. To follow-up on this finding, the number of correctly identified real and made-up items (measured by the multiple-choice questionnaire in Appendix B) were included as predictors of the individual differences at the R stage. Not surprisingly, the knowledge measures were highly significant ($\chi^2 = 149.5$, $\mathrm{df} = 2$), showing that variability at the R stage is strongly related to knowledge differences. The regression effects for real and fake items were estimated as 0.79 (0.07) and 0.19 (0.05), respectively, suggesting that knowledge differences for real items were more predictive of individual retrieval differences for real than for fake items. This result is similar to the one obtained for the NfC scores and strongly suggests that there is a qualitative difference between knowing an item and knowing that an item is made up. In fact, the positive effect of NfC and knowledge for real items and the corresponding null or weaker effects for made-up items suggest that the real and made-up items may be more appropriately represented by a two- than a one-dimensional R stage that captures the distinctive nature of the two item types.

3.3.4. Conclusion The results obtained under the RES model provide a parsimonious and readily interpretable representation of the data. Most importantly, we found that the RES model could account for overclaiming, the effects of the covariates, as well as the experimental manipulations. The parameters estimated for the three postulated response stages yielded detailed information about the overclaiming process. According to the RES model, overclaiming is caused mostly by an increase in a person's tendency to edit the initial familiarity assessment. However, the results of the Power condition also point to a second mechanism: Respondents primed with power arrived at a higher initial familiarity assessment than respondents in the Control condition. The positive relationship between the first and second mechanisms can account for the finding that respondents who were more familiar with the domain were also more likely to overclaim in the Power condition. Thus, in this case, overclaiming does not seem to be a simple matter of trying to compensate for a lack of knowledge, which implies a negative relationship between editing tendencies and higher familiarity levels. Instead, we also found the opposite pattern of respondents editing their responses even more when they experienced higher confidence in their initial familiarity assessments.

Tendencies to overclaim are malleable and can vary across items, but they are also person-specific and can be related to individual-difference measures. We found that NfC scores provided useful insights about sources of variability in the initial-familiarity assessments and that SDE scores predicted tendencies to edit. But, most importantly, we also found that by analyzing each response stage separately we can diagnose and adjust for editing effects using the RES framework. The strong association between the number of correct answers and the individual-difference effects estimated at the R stage suggests that the item parameters at this stage may provide a more accurate representation of the respondents' actual familiarity levels than the self-reports. Thus, if the RES model provides a satisfactory fit of the data, it could prove useful in

correcting for socially desirable responding, even if external information about the true response is lacking.

4. Discussion

Asking sensitive or personal questions can both lower response rates and increase item non-response and misreports. Although non-response is easily diagnosed, misreports are not. There is some evidence to suggest that misreporting results from a motivated process in which respondents edit their answers before they report them (Tourangeau, Rips, & Rasinski, 2000). Specifically, respondents tend to overreport socially desirable behaviors and underreport less socially desirable ones. Because these effects of response editing move in the same direction, they systematically bias survey estimates. To address this issue, we proposed the RES model, which takes into account that respondents may arrive at their answers via multiple response stages.

The RES framework assumes that when respondents are asked questions about personal or sensitive issues, they may want to give honest answers; but, after a moment of deliberation, they may also want to present themselves in a favorable light with the result that items may measure both the actual behaviors of the respondents as well as the respondents' tendency to edit their responses. Allowing for both response processes may lead to more-valid conclusions about item characteristics and drivers of individual differences. For example, survey methods (such as whether the interview is conducted face-to-face or via an online questionnaire) may affect the degree to which respondents edit their responses but not the actual behavior under study. Relating covariates to the hypothesized response processes or direct experimental manipulations of the response stages can prove critical for understanding the determinants of the responses. These RES model features become increasingly important as we learn more about factors that influence the degree to which respondents misreport.

The application showed that a power prime can trigger respondents to express that they know more and that they feel more certain about their knowledge. This result is consistent with much recent work showing that socially desirable responding in surveys is largely contextual and depends both on the respondents' situation and on features of the data-collection situation (Tourangeau & Yan, 2007). For example, recent work by John, Acquisti, and Loewenstein (2011) demonstrates that a person's willingness to divulge personal information is reduced in online surveys when a privacy policy is displayed. A possible interpretation of this finding is that reading a privacy policy can remind respondents to protect their privacy which, in turn, can lead them to edit their answers. In addition to privacy and self-presentation concerns, it has also been shown that respondents edit their responses to screening questions when they expect an affirmative response to lead to multiple follow-up questions that are tedious and time-consuming (Yang et al., 2010). Clearly, this list is by no means exhaustive and more research is needed to understand the conditions that both facilitate and inhibit truthful responding (Mazar & Ariely, 2006).

One immediate avenue for future research is to combine randomized-response models with the RES framework presented here. Although the randomized-response technique has been shown to increase truthful answering, recent applications also demonstrated that they do not eliminate response bias (Böckenholt & van der Heijden, 2007). Perhaps for reasons that are similar to the finding reported by John et al. (2011), some respondents may become self-protective when instructed about the randomized-response design. The RES framework could be used to detect this response behavior and to investigate the extent to which respondents distinguish and react positively to the different levels of privacy protection offered by randomized-response methods.

In conclusion, not all self-reports can be taken as unbiased, candid, and accurate. Respondents who consider questions as private or relevant for themselves may choose to edit their answers, which—when not corrected—leads to systematically biased estimates. By taking into

TABLE 5.
Simulation results of RES model with three item categories.

Effects	Pop. value	$n = 5{,}000$			$n = 1{,}000$			$n = 500$		
		Mean Est.	$\overline{\text{SE}}$	SD/$\overline{\text{SE}}$	Mean Est.	$\overline{\text{SE}}$	SD/$\overline{\text{SE}}$	Mean Est.	$\overline{\text{SE}}$	SD/$\overline{\text{SE}}$
$\gamma_1^{(R)}$	−1.000	−0.999	0.095	1.000	−1.005	0.209	1.129	−1.017	0.282	1.188
$\gamma_2^{(R)}$	−0.500	0.500	0.080	1.000	−0.500	0.179	1.128	−0.500	0.242	1.252
$\gamma_3^{(R)}$	0.000	0.000	0.082	0.988	0.017	0.173	1.143	0.011	0.228	1.329
$\gamma_4^{(R)}$	0.500	0.498	0.088	1.000	0.514	0.181	1.193	0.504	0.227	1.374
$\gamma_5^{(R)}$	1.000	1.012	0.104	1.096	1.025	0.196	1.206	1.022	0.238	1.378
$\gamma_1^{(E)}$	−0.300	−0.303	0.059	1.000	−0.306	0.137	1.117	−0.321	0.196	1.321
$\gamma_2^{(E)}$	−0.150	−0.154	0.076	0.974	−0.156	0.176	1.188	−0.158	0.241	1.261
$\gamma_3^{(E)}$	0.000	−0.001	0.104	0.990	−0.027	0.236	1.190	−0.011	0.307	1.264
$\gamma_4^{(E)}$	0.150	0.150	0.136	1.055	0.123	0.302	1.222	0.179	0.363	1.253
$\gamma_5^{(E)}$	0.300	0.271	0.193	1.122	0.238	0.391	1.232	0.338	0.447	1.221
σ_1	0.500	0.502	0.097	0.963	0.531	0.174	0.977	0.592	0.264	0.894
σ_2	0.500	0.484	0.129	1.062	0.562	0.359	0.827	0.676	0.664	0.910
τ_2	1.000	0.999	0.053	1.000	0.988	0.135	1.235	0.971	0.188	1.300
ω_2	0.500	0.508	0.122	1.000	0.533	0.287	1.188	0.514	0.365	1.214

account the different stages of the response process, the RES model may prove to be a useful tool to improve the quality of survey data.

Acknowledgements

This research was supported in part by grants from the Social Sciences and Humanities Research Council of Canada and the Canadian Foundation of Innovation.

Appendix A. Simulation Studies

To assess the estimation bias of the RES model, a number of simulation studies were performed. Here, we present the results of an RES model with five items having either three or four response categories, respectively.

A.1. RES Model with Three Response Categories

The parameter values for the RES model with three response categories are reported in Table 5. The random effects $\theta_i^{(R)}$, $\theta_i^{(E)}$, and $\theta_i^{(S)}$ were specified to be equally correlated with

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1 + \sigma_2 & \sigma_2 & \sigma_2 \\ \sigma_2 & \sigma_1 + \sigma_2 & \sigma_2 \\ \sigma_2 & \sigma_2 & \sigma_1 + \sigma_2 \end{pmatrix}, \tag{A.1}$$

and $\sigma_1 = \sigma_2 = 0.5$. Table 5 summarizes the estimation results for the three sample sizes $n = 5{,}000$, $n = 1{,}000$ and $n = 500$ based on 500 replications each. We report the estimated mean parameter values, the mean standard error, as well as the ratio of the mean standard error and

TABLE 6.
Simulation results of RES model with four item categories.

Effects	Pop. value	$n = 1{,}000$			$n = 500$		
		Mean Est.	$\overline{\text{SE}}$	SD/$\overline{\text{SE}}$	Mean Est.	$\overline{\text{SE}}$	SD/$\overline{\text{SE}}$
$\gamma_1^{(R)}$	1.000	1.108	0.320	1.079	0.885	0.324	1.381
$\gamma_2^{(R)}$	0.500	0.549	0.218	1.143	0.468	0.234	1.283
$\gamma_3^{(R)}$	0.000	0.009	0.149	1.036	0.032	0.157	1.118
$\gamma_4^{(R)}$	0.500	0.554	0.218	1.153	0.453	0.227	1.259
$\gamma_5^{(R)}$	1.000	1.119	0.321	1.125	0.921	0.335	1.348
$\gamma_1^{(E)}$	0.300	0.294	0.281	1.068	0.272	0.290	1.178
$\gamma_2^{(E)}$	0.150	0.097	0.343	1.076	0.115	0.344	1.226
$\gamma_3^{(E)}$	0.000	−0.088	0.405	1.054	−0.077	0.407	1.030
$\gamma_4^{(E)}$	0.150	0.098	0.339	1.122	0.101	0.350	1.263
$\gamma_5^{(E)}$	0.300	0.295	0.280	1.059	0.216	0.278	1.178
φ_1	0.500	0.535	0.127	1.057	0.430	0.141	1.106
φ_2	−0.500	−0.530	0.149	1.085	−0.451	0.181	1.028
λ_{11}	1.000	1.149	0.362	0.998	0.891	0.389	1.231
λ_{21}	0.500	0.418	0.316	1.128	0.345	0.347	1.188
λ_{22}	0.866	0.819	0.304	1.175	0.606	0.360	1.203
λ_{31}	0.500	0.476	0.339	1.083	0.371	0.414	1.226
λ_{32}	0.289	0.295	0.534	1.036	0.358	0.561	1.280
λ_{32}	0.816	0.630	0.407	0.878	0.576	0.380	1.182
τ_2	−1.386	−1.400	0.361	0.973	−1.283	0.418	1.103
τ_3	−0.693	−0.661	0.122	1.011	−0.609	0.150	1.130
ω_2	−0.250	−0.268	0.381	1.132	−0.238	0.391	1.205
ω_3	−0.500	−0.519	0.396	1.111	−0.435	0.390	1.268

standard deviation of the estimated parameter values. For $n = 5{,}000$, the estimated bias is small and the mean standard errors agree well with the standard deviations of the estimated parameters. For the smaller sample sizes $n = 1{,}000$ and $n = 500$, the bias of the item parameters continues to be small but the bias in the standard errors increases. They appear to be systematically smaller than the standard deviations of the estimated parameter values. For each of the fitted models, we also computed the expected a posteriori (EAP) person scores. The results of these analyses are reported in the section "Recovery of Item and Person Parameters".

A.2. RES Model with Four Response Categories

The setup of the RES model with four response categories differed from the previous simulation study in two ways. First, covariates were included at the R and E stages of the model. Second, although the elements of the covariance matrix were of equal size as in the previous simulation study, their estimation was unconstrained. Specifically, we set $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}$ and estimated the corresponding elements of the Cholesky matrix of $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$ with $\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & \lambda_{22} & 0 \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{pmatrix}$.

The first two columns of Table 6 list the effects and the chosen parameter values for five items for both the response-formation and editing stages, the elements of the Cholesky matrix,

as well as the threshold values of the response-formation stage and the category attractiveness values of the editing stage. Four groups are specified that differ in the item parameters for the R and E stages of the model. Specifically, for Group 1 the item effects are $\boldsymbol{\gamma}^{(R)} = (1, 0.5, 0, 0.5, 1)$ and $\boldsymbol{\gamma}^{(E)} = (0.3, 0.15, 0, 0.15, 0.3)$. The corresponding item effects for Group 2 are $\boldsymbol{\gamma}^{(R)}$ and $\boldsymbol{\gamma}^{(E)} + \varphi_2$, for Group 3, $\boldsymbol{\gamma}^{(R)} + \varphi_1$ and $\boldsymbol{\gamma}^{(E)}$ and, for Group 4, $\boldsymbol{\gamma}^{(R)} + \varphi_1$ and $\boldsymbol{\gamma}^{(E)} + \varphi_2$, where $\varphi_1 = 0.5$ and $\varphi_2 = -0.5$. The sample sizes of the four groups were specified to be equal. The remaining columns of Table 6 report the estimated parameters, mean standard errors, as well as the ratio of the mean standard error and standard deviation of the estimated parameter values for the two sample sizes $n = 500$ and $n = 1{,}000$. These values were obtained based on 500 replications. As in the previous simulation study, we find that the estimation bias is small for both sample sizes and that for $n = 500$, the standard errors appear systematically smaller compared to the standard deviations of the estimated parameter values. Likelihood-ratio tests may provide more accurate inferences at this sample size.

Appendix B. Item Questionnaire

1. Sciatica is:
   - an anxiety-reducing drug
   - caused by the compression of nerves
   - a hormone
   - a protein
   - none of the above
2. Meiosis is:
   - a chromosome
   - a hormone
   - a type of cell division
   - a skin disease
   - none of the above
3. Antigen is:
   - a hormone
   - a protein
   - a disease
   - a virus
   - none of the above
4. Meta-toxins are:
   - produced by cancer cells
   - pain relievers
   - chemical agents
   - used to develop vaccines
   - none of the above
5. Bio-sexual
   - refers to the reproduction of plants
   - refers to non-chemical birth-control methods
   - refers to the passion for biology
   - refers to an account of someone's sexual life

- none of the above

6. Retroplex is:

   - a part of cell structures
   - a neck muscle
   - an involuntary movement
   - the inability to recall past events
   - none of the above

References

Benitez-Silva, H., Buchinsky, M., Chan, H.-M., Cheidvasser, S., & Rust, J. (2004). How large is the bias in self-reported disability? *Journal of Applied Econometrics*, *19*, 649–670.

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.

Bound, J., Brown, C.C., & Mathiowetz, N. (2001). Measurement error in survey data. In E.E. Learner & J.J. Heckman (Eds.), *Handbook of econometrics* (pp. 3705–3843). Amsterdam: North-Holland.

Bowman, D., Heilman, C., & Seetharaman, P. (2004). Determinants of product-use compliance behavior. *Journal of Marketing Research*, *41*, 324–338.

Bradlow, E.T., & Zaslavsky, A.M. (1999). A hierarchical latent variable model for ordinal data from a customer satisfaction survey with "no answer" responses. *Journal of the American Statistical Association*, *94*, 43–52.

Brinõl, P., Petty, R.E., Valle, C., Rucker, D.D., & Becerra, A. (2007). The effects of message recipients' power before and after persuasion: a self-validation analysis. *Journal of Personality and Social Psychology*, *93*, 1040–1053.

Cacioppo, J.T., & Petty, R.E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131.

Cacioppo, J.T., Petty, R.E., Feinstein, J.A., & Jarvis, W.B.G. (1996). Dispositional differences in cognitive motivation: the life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*, 197–253.

Cacioppo, J.T., Petty, R.E., & Kao, C.F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*, 306–307.

Campbell, W.K., Goodie, A.S., & Foster, J.D. (2004). Narcism, confidence, and risk attitude. *Journal of Behavioral Decision Making*, *17*, 297–311.

Galinsky, A.D., Gruenfeld, D.H., & Magee, J.C. (2003). From power to action. *Journal of Personality and Social Psychology*, *85*, 453–466.

Gill, P., Murray, W., & Wright, M. (1981). *Practical optimization*. San Diego: Academic Press.

Harvey, J.W., & McCrohan, K. (1988). Voluntary compliance and the effectiveness of public and non-profit institutions: American philanthropy and taxation. *Journal of Economic Psychology*, *9*, 369–386.

Hewitt, P.L., Flett, G.L., Sherry, S.B., Habke, M., Parkin, M., Lam, R.W., McMurtry, B., Ediger, E., Fairlie, P., & Stein, M.B. (2003). The interpersonal expression of perfection: perfectionistic self-presentation and psychological distress. *Journal of Personality and Social Psychology*, *84*, 1303–1325.

Holtgraves, T. (2004). Social desirability and self-reports: testing models of socially desirable responding. *Personality & Social Psychology Bulletin*, *30*, 161–172.

Hsiao, C., Sun, B.-H., & Morwitz, V.G. (2002). The role of stated intentions in new product purchase forecasting. *Advances in Econometrics*, *16*, 11–28.

John, L.K., Acquisti, A., & Loewenstein, G. (2011). Strangers on a plane: context-dependent willingness to divulge sensitive information. *Journal of Consumer Research*, *37*, 858–873.

Johnson, T.R., & Bolt, D.M. (2010). On the use of factor-analytic multinomial logit item response models to account for individual differences in response style. *Journal of Educational and Behavioral Statistics*, *35*, 92–114.

Magee, J.C., & Galinsky, A.D. (1992). Social hierarchy: the self-reinforcing nature of power and status. *Academy of Management Annals*, *2*, 351–398.

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

Mittal, V., & Kamakura, W. (2001). Satisfaction, repurchase intent, and repurchase behavior: investigating the moderating effect of customer characteristics. *Journal of Marketing Research*, *38*, 131–142.

Öhman, N. (2011). Buying or lying - the role of social pressure and temporal disjunction of intention assessment and behavior on the predictive ability of good intentions. *Journal of Retailing and Consumer Services*, *18*, 194–199.

Orlando, M., & Thissen, D. (2000). New item fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, *24*, 50–64.

Paulhus, D.L. (2002). Socially desirable responding: the evolution of a construct. In H. Braun, D.N. Jackson, & D.E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 67–88). Hillsdale: Erlbaum.

Paulhus, D.L., Harms, P.D., Bruce, M.N., & Lysy, D.C. (2003). The over-claiming technique: measuring bias independent of accuracy. *Journal of Personality and Social Psychology*, *84*, 681–693.

Reingen, P. (1978). On inducing compliance with requests. *Journal of Consumer Research*, *5*, 96–102.

Rorer, L.G. (1965). The great response-style myth. *Psychological Bulletin*, *63*, 129–156.

Sadowski, C.J., & Gülgöz, S. (1992). Internal consistency and test-retest reliability of the need for cognition scale. *Perceptual and Motor Skills*, *74*, 610.

Samejima, F. (1997). Graded response model. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85–100). Berlin: Springer.
Simon, A.F., Fagley, N.S., & Halleran, J.G. (2004). Decision framing: moderating effects of individual differences and cognitive processing. *Journal of Behavioral Decision Making*, *17*, 77–93.
Sinha, R.K., & Mandel, N. (2008). Preventing music piracy: the carrot or the stick? *Journal of Marketing*, *72*, 1–15.
Swets, J.A. (1964). *Signal detection and recognition by human observers*. New York: Wiley.
Tellis, G.J., & Chandrasekaran, D. (2010). Extent and impact of response biases in cross-national survey research. *International Journal of Research in Marketing*, *27*, 329–341.
Toma, C.L., Hancock, J., & Ellison, N. (2008). Separating fact from fiction: an examination of deceptive self-presentation in online dating profiles. *Personality & Social Psychology Bulletin*, *34*, 1023–1036.
Tourangeau, R., Rips, L.J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.
Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*, 859–883.
van Soest, A., & Hurd, M. (2008). A test for anchoring and yea-saying in experimental consumption data. *Journal of the American Statistical Association*, *103*, 126–136.
Wirtz, J., & Kum, D. (2004). Consumer cheating on service guarantees. *Journal of the Academy of Marketing Science*, *32*, 159–175.
Wlaczyk, J., Schwartz, J.P., Clifton, R., Adams, B., Wei, M., & Zha, P. (2005). Lying person-to-person about life events: a cognitive framework for lie detection. *Personnel Psychology*, *58*, 141–170.
Wosinska, M. (2005). Direct-to-consumer advertising and drug therapy compliance. *Journal of Marketing Research*, *42*, 323–332.
Yang, S., Zhao, Y., & Dhar, R. (2010). Modeling the underreporting bias in panel survey data. *Marketing Science*, *29*, 525–539.

Manuscript Received: 13 JUL 2011
Published Online Date: 3 DEC 2013