MARTIJN G. DE JONG, RIK PIETERS, and JEAN-PAUL FOX*

The authors present a polytomous item randomized response model to measure socially sensitive consumer behavior. It complements established methods in marketing to correct for social desirability bias a posteriori and traditional randomized response models to prevent social desirability bias a priori. The model allows for individual-level inferences at the construct level while protecting the privacy of respondents at the item level. In addition, it is possible to incorporate covariates into various parts of the model. The proposed method is especially useful to study social issues in marketing. In the empirical application, the authors use a two-group experimental survey design and find that with the new procedure, participants report their sensitive desires more truthfully, with significant differences between socioeconomic groups. In addition, the method performs better than methods based on social desirability scales. Finally, the authors discuss truthfulness in data collection and confidentiality in data utilization.

Keywords: randomized response, item response theory, item randomized response, social desirability, response styles

Reducing Social Desirability Bias Through Item Randomized Response: An Application to Measure Underreported Desires

Ever since early articles by Churchill (1979) and Anderson and Gerbing (1988), marketing has developed systematic and rigorous approaches to quantify latent constructs in surveys. Such latent constructs in marketing include the goals, attitudes, desires, emotions, and intentions of consumers, managers, and firms, which are intrinsically unobservable and for which self-reports are the prime source of information. However, especially for more sensitive topics, these data are often contaminated by socially desirable responding (SDR; Mick 1996)—that is, participants' tendency to describe themselves in favorable terms by adhering to socioculturally sanctioned norms. Socially desirable responding has been recognized as a serious problem that can adversely affect the validity of studies in many social science disciplines. To cope with this, various approaches have been proposed to control for SDR bias in marketing research.

First, there are attempts to check for social desirability bias after the data have been collected using dedicated SDR scales, such as the Marlowe–Crowne scale or Paulhus's (2002) balanced inventory of desirable responding (for applications in marketing, see Mick 1996; Podsakoff et al. 2003; Steenkamp, De Jong, and Baumgartner 2010). However, the use of SDR scales has been contested (Paulhus 2002; Smith and Ellingson 2002). The primary reason is the difficulty of separating valid personality content in the SDR measures from the bias they are meant to measure (Paulhus 2002; Steenkamp, De Jong, and Baumgartner 2010). In addition, SDR scales are general traitlike measures and may have small effects on domain-specific surveys.

Second, there are various approaches to prevent SDR from biasing the measures in the first place, such as by using indirect questioning and bogus pipeline techniques (e.g., Aguinis, Pierce, and Quigley 1993; Fischer 1993; Hill, Dill, and Davenport 1988; Jones and Sigall 1971; Roese and Jamieson 1993; Tourangeau and Smith

1996). Although they may partially alleviate SDR, the effectiveness of these techniques is limited because they may introduce other biases (indirect questions), and their implementation may be expensive and prone to ethical issues (bogus pipeline) (Tourangeau, Rips, and Rasinski 2000).

Therefore, randomized response (RR) methodologies have been proposed to prevent SDR effectively (Clark and Desharnais 1998; Lamb and Stem 1978; Lensvelt-Mulders et al. 2005; Reinmuth and Geurts 1975; Warner 1965). These methodologies aim to prevent SDR bias during data collection by providing privacy protection through a randomization mechanism, after which statistical techniques are used to infer the true responses of the participants on the measures. However, a major drawback of RR models to date is that only aggregate-level inferences can be obtained, but no individual-level inferences. This prevents insights into the possible determinants and consequences of the sensitive construct under study and is an explanation why RR methods, despite their introduction into the marketing literature, have not been widely applied yet (Lamb and Stem 1978; Reinmuth and Geurts 1975).

Here, we propose an item RR model that does not suffer from this limitation. The model combines RR techniques (Fox and Tracy 1986; Lensvelt-Mulders et al. 2005; Warner 1965) with item response theory (IRT) (Lord and Novick 1968). The essential idea is to provide privacy protection to survey participants at the level of the individual items in multi-item measures. However, by relating multiple inter-correlated items, it is still possible to make individual-level inferences at the level of the latent construct (Böckenholt and Van der Heijden 2007; Fox 2005).

We contribute to the literature in the following ways: First, to the best of our knowledge, the current study is the first to introduce item RR models to marketing, in which socially sensitive responding is commonly at stake. Second, whereas existing item RR models only accommodate dichotomous items, the proposed model is designed for polytomous items. Because items in marketing research typically have five-point or seven-point ordinal response formats (Bearden and Netemeyer 1999; Brunel and Hensel 1992 [and subsequent volumes]), an extension to polytomous item formats is crucial. In addition, our proposed model is fully explanatory (De Boeck and Wilson 2004) by allowing for the inclusion of covariates in both the structural model and the measurement model. Jointly, this enables individual-level inferences about the antecedents and consequences of the construct of interest, as well as explanations of the determinants of the specific item parameters. Another distinguishing aspect of the model is that we allow for a distinct group of respondents who do not follow the instructions and thus would bias the findings if they remain undetected. Finally, we provide an easy method to check model fit. We compare the performance of the item RR model with the yardstick approach in marketing—namely, the use of SDR scales.

Our method is especially suited to survey social issues in marketing (e.g., responsible behavior with respect to public health issues, true behavior with respect to environmental issues) and to measure sensitive issues in transformative consumer behavior (e.g., tobacco, alcohol, and other drug consumption; gambling; financial and medical decision making). In these domains, threats to consumers can be either extrinsic, if certain responses carry the risk of legal sanctions (e.g., possession of black money), or intrinsic, if the questions pertain to topics that are personal to the respondents.

Within this domain of research, our empirical application pertains to the economically and socioculturally important but virtually uncharted domain of adult entertainment (Cronin and Davenport 2001), focusing on consumers' desires for products and services in this domain. Although the adult entertainment industry is one of the largest and most profitable industries worldwide,[1] little is known about individual-level consumption and desires for the category. Not only is research on the topic scarce because of its low regard in marketing and the social sciences at large (Cronin and Davenport 2001), but (crucially) the consumption of adult entertainment is also perceived as almost "immeasurable" by organizations such as ACNielsen. In general, society stigmatizes people who openly admit such desires, so survey participants tend to be hesitant to provide honest answers to direct questions that probe this taboo.

RR MODELS

When responses to questionnaires are influenced by SDR, people consciously provide untruthful, distorted answers to present themselves in a better light or to prevent threats to image and self-esteem, and these response tendencies harm measurement validity. Because RR techniques can prevent SDR from biasing survey research (Lensvelt-Mulders et al. 2005) and because recent advances in modeling the data from RR techniques (Böckenholt and Van der Heijden 2007; Fox 2005) hold considerable promise for marketing, we discuss these techniques more extensively here. We begin by illustrating the aggregate RR model, starting with a description of the aggregate RR model for dichotomous survey items, because this is the way the method has traditionally been used and it is the most straightforward to understand. Subsequently, we present the RR model for polytomous items, which is most relevant for marketing (Bearden and Netemeyer 1999).

Univariate Dichotomous Data

There are two main classes of RR techniques: the related question method (RQM), also known as Warner's (1965) method, and the unrelated question method (UQM; Greenberg et al. 1969). To describe the procedures briefly, suppose that a researcher wants to estimate the proportion of people who buy a certain sensitive product X. In the RQM, people are required to respond "yes" or "no" to one of two statements: (1) "I buy product X," or (2) "I do not buy product X." The interviewees select one of the two statements with probabilities p and $(1-p)$ using a random device (e.g., a coin, dice, cards), but they do not reveal their true answers. In the UQM, the second statement is not related to the first and is completely innocuous (e.g., whether a person's birthday is before a certain day of the year). Greenberg and colleagues (1969) show that the UQM is more efficient than the RQM without needing to reveal the true individual answers.

---

[1] See http://internet-filter-review.toptenreviews.com/internet-pornography-statistics.html (last accessed November 2008).

There are two well-known modifications of the UQM that are statistically equivalent. In both variations, the population prevalence of the unrelated (innocuous) question is known. This is possible by using an innocuous question with known prevalence or by using the forced response method. In the forced response method, participants are asked to answer the sensitive question (with probability p_1), or a forced response is given (with probability $1 - p_1$). If a forced response is given, a random device determines whether a forced "yes" response is given (with probability p_2) or whether a forced "no" response is given (with probability $1 - p_2$). In this case, the forced responses represent answers to an unrelated question with known population prevalence. The forced response method is known to be one of the most efficient RR designs, and it is easily implemented (Fox 2005). For these reasons, we use it in our proposed methodology. The other modified version of the UQM requires (multiple) innocuous questions with known population prevalence, which may be difficult to obtain in practice.

A simple example provides the idea behind the forced response method. People are asked whether they buy a certain sensitive product, using a simple yes/no response format (dichotomous). In addition, participants flip a coin. The interviewer does not observe the coin flip. People are instructed to answer the question truthfully if the coin comes up heads. Otherwise, they can ignore the question and just say "yes," no matter what their true opinion would have been. Thus, the coin flip serves as a mechanism to protect the participant. This simple setup provides full privacy protection because the interviewer (who does not see the coin flip) cannot know whether a "yes" means that a person buys the product or whether the coin came up tails.

Despite the impossibility of knowing the true answers of individual participants, it is still possible to estimate the proportion of the sample who would privately admit to buy the product. Let π denote the unknown proportion who has actually bought the product, and let λ the observed proportion of "yes" responses. It follows that $\lambda = (\pi/2) + (1/2)$, and the reason for this is straightforward: It is known that participants who flipped heads and have bought such products answered "yes" and that participants who flipped tails are instructed to answer "yes" regardless of whether they are engaged in the behavior. As a result, $\pi = 2\lambda - 1$.

Although RQM and UQM and their modifications allow for aggregate statements on the sample level, nothing can be said about individual-level responses, and thus no individual-level inferences can be made, such as relating the answers to potential determinants or consequences. Furthermore, the techniques are restricted to univariate binary response data, and they do not allow a hierarchical analysis of the RR data when, for example, participants are nested in groups and responses from members of one group are likely to be correlated.

Polytomous Data

Randomized response techniques have been developed mostly for dichotomous response formats, but this format is rarely used in marketing. The majority of items (questions) in marketing surveys use five-point or seven-point ordinal response formats (Bearden and Netemeyer
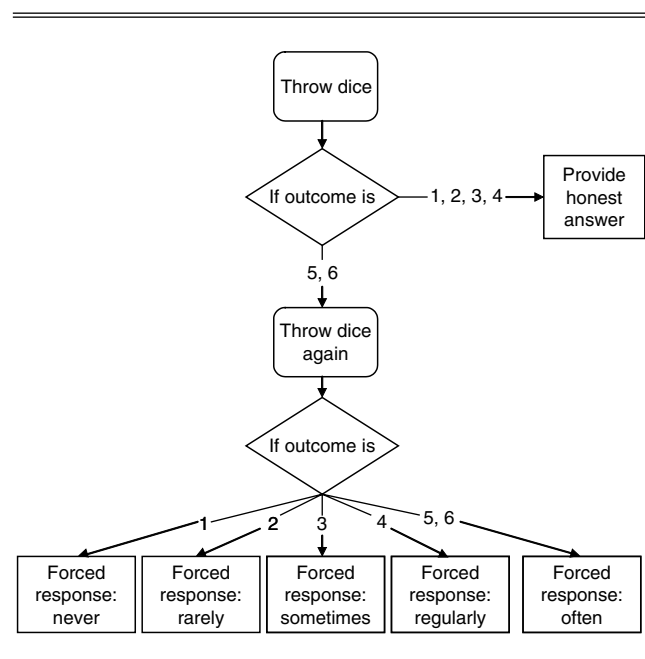
1999). For such cases, a multiproportions version of the RR technique can be applied. In particular, the more efficient forced response method is easily modified to more than two response categories—for example, $c = 1, \ldots, C$ response categories.

To illustrate, suppose that a question is asked about the frequency with which consumers have certain desires, with a five-point response format: "never," "rarely," "sometimes," "regularly," and "often." Next, a randomization device could be used so that with a known probability p_1, a true response should be given, and with probability $1 - p_1$, a forced choice should be given. Conditional on the requirement to give a forced choice, the probability of a certain response option is $p_{2,c}$. One way to operationalize the randomization process for such polytomous items is to use dice throws by consumers, but coin tosses or other randomization devices could be used as well. Figure 1 presents a flowchart for such a procedure, which forms the basis of our subsequent empirical illustration. The probability of obtaining "often" on the scale is higher than the probability of obtaining "never" to "regularly" because both the outcomes 5 and 6 of the die are mapped onto response option "often," which is readily accommodated statistically. We define the probability λ of response c ($c = 1, \ldots, C$) as follows:

$$(1) \qquad \lambda_c = P(Y = c) = p_1 \pi_c + (1 - p_1) p_{2,c},$$

where p_1 is the probability of being required to answer honestly (in our example, $p_1 = 4/6$), π_c is the true probability of response c when directly and honestly answering the question, and $p_{2,c}$ is the probability of some forced response c given that a forced response must be provided. For the die, the probability $p_{2,c}$ is equal for the response options

Figure 1
ILLUSTRATING RR TECHNIQUES FOR ORDINAL ITEMS BASED ON DICE THROWS

"never," "rarely," "sometimes," and "regularly" and is different for response option "often." The likelihood of the data is then given by a multinomial likelihood:

$$(2) \quad L(\pi_1, \ldots, \pi_C; Y) = \frac{n!}{n_1! n_2! \cdots n_C!} \lambda_1^{n_1} \lambda_2^{n_2} \cdots \lambda_C^{n_C} = n! \prod_{c=1}^{C} \frac{\lambda_i^{n_i}}{n_i!},$$

where n_c is the number of participants in cell c. Because p_1 and p_2 are known, being set by the researcher, it is easy to obtain the parameters π_c, framed within the λ_c. Note again that though this method allows for aggregate sample-level statements, in this case of polytomous items, nothing can be said about individual-level responses. Moreover, the method can only handle nonnested univariate response data.

ITEM RR MODEL

Originally developed for single items, RR techniques can also be used for multiple, correlated items, and this is important. Item response theory models for polytomous data can place participants on a continuum that represents their true score on some latent variable (construct) based on multiple RR items. Our proposed model uses IRT to make individual-level inferences from multiple correlated RR items. We first discuss the most popular IRT model for polytomous data.

Samejima's IRT Model

Item response theory models have been developed to relate a set of either dichotomous or polytomous items to an underlying latent construct. Such models are nonlinear in nature and have several advantages over classical test theory (Churchill 1979), such as item-free calibration, sample-independent item parameters, incorporation of the ordinal nature of the data, separation of item and person parameters, incorporation of floor and ceiling effects, enabling of complex sampling designs, and incorporation of missing data processes (see De Jong, Steenkamp, and Fox 2007; De Jong et al. 2008).

We model the ordinal data that typically arise in marketing from Likert, semantic differential, and similar items based on the work of Samejima (1969). The model specifies the operating characteristic $P_c(\xi_i)$ of an item response—that is, the probability of a particular response c as a function of a latent construct ξ_i. We define the operating characteristic for a particular item as follows:

$$(3) \quad P_c(\xi_i) = \frac{1}{\sqrt{2\pi}} \int_{a[\xi_i - \gamma_c]}^{a[\xi_i - \gamma_{c-1}]} \exp\left(-\frac{1}{2}t^2\right) dt,$$

where the parameter a is the "discrimination" parameter and the parameter γ is a "threshold" parameter, both of which can vary across items. The thresholds govern the frequency with which items are endorsed and obey the following order restriction: $\gamma_1 < \ldots < \gamma_C$. The discrimination parameters indicate how well items reflect ("load on") the latent construct. Indexing the operating characteristic by k, we have the normal ogive variant of Samejima's graded response model given by the following:

$$(4) \quad P(\tilde{Y}_{ik} = c \mid \xi_i; a_k, \gamma_{k,c-1}, \gamma_{k,c})$$
$$= \frac{1}{\sqrt{2\pi}} \int_{a_k[\xi_i - \gamma_{k,c}]}^{a_k[\xi_i - \gamma_{k,c-1}]} \exp\left(-\frac{1}{2}t^2\right) dt,$$

where $\tilde{Y}_{ik}$ is a latent response (i.e., it is latent because the true response is randomized before it is observed) and ξ_i is the latent score of individual i on the latent construct. This model specifies the conditional probability that participant i provides response category c (c = 1, ..., C) for item k. The thresholds $\gamma_{k,c}$ are measured on the same scale as ξ_i and determine the difficulty of responding above a certain category c. The threshold $\gamma_{k,c}$ is theoretically defined as the value on the latent scale, so the probability of responding above a value c is .5, for c = 1, ..., C − 1.

A Polytomous Item RR Model

Aggregate RR models thwart individual-level inference because they focus on single items only. However, such inferences become possible by considering multiple items that are each influenced by ξ_i and that are each measured by RR techniques, and this idea is the basis of the proposed polytomous item RR model. We formulate our basic model as follows:

$$(5) \quad \begin{array}{l} P(Y_{ik} = c) = p_1 \pi_{ikc} + (1 - p_1) p_{2,c} \\ \pi_{ikc} = P(\tilde{Y}_{ik} = c \mid \xi_i; a_k, \gamma_{k,c-1}, \gamma_{k,c}). \end{array}$$

The first part of the equation specifies the RR component, and the term π_{ikc} captures the IRT part of the model (the parameter π_{ikc} depends on the response category, item, and participant). Note the difference between $\tilde{Y}_{ik}$ and Y_{ik}: The former is the latent response and is unobserved because it is randomized before observation. The RR process influences Y_{ik} but not $\tilde{Y}_{ik}$. The complete response process for a five-point ordinal item k appears in Figure 2 (the parts of the tree resulting from π_{ik4} and π_{ik5} are not displayed).

Consider the upper part of the tree. When confronted with an item k, individual i answers $\tilde{Y}_{ik} = 1$ with probability π_{ik1}. This choice is observed only with probability p_1 because with probability $1 - p_1$, a forced response must be given because of the RR procedure. When a forced response must be given under the RR mechanism, the probability of each response is $p_{2,c}$. The interpretation of the rest of the tree follows the same logic.

An IRT analysis of RR data is an important extension of traditional RR techniques, and there are several notable advantages: (1) It enables the separation of item characteristic parameters (difficulty and discrimination) and person characteristic parameters (trait levels), (2) different sets of items can be used to measure people on one common scale, (3) it handles measurement error at the individual level in a natural way, and (4) a robust trait estimate is obtained.[2]
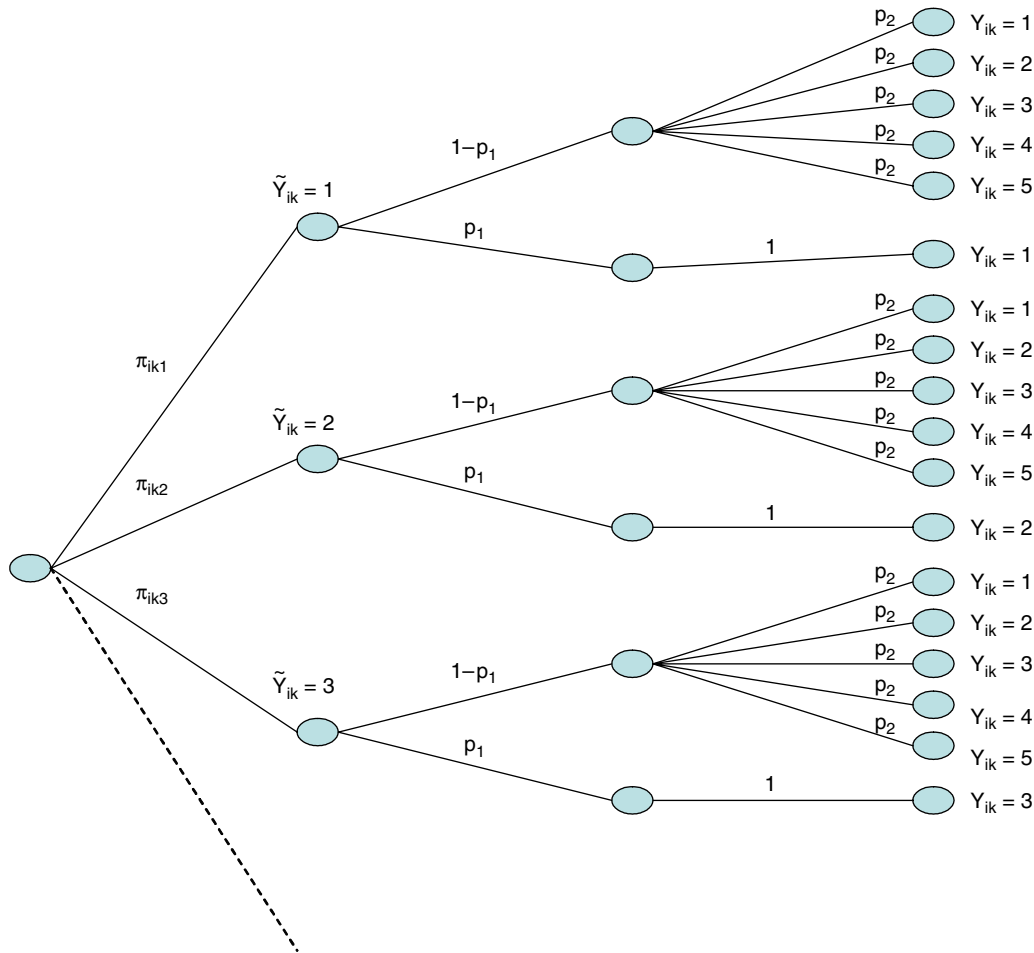
We can extend the basic model in Equation 5 in straightforward ways. For example, if antecedents of the latent trait are of interest, a structural model can simultaneously be imposed on top of it:

$$(6) \quad \xi_i = \mathbf{X}_i' \boldsymbol{\beta} + \varepsilon_i,$$

where $\mathbf{X}_i$ contains K covariates, $\boldsymbol{\beta}$ is a K-dimensional parameter vector, and $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ is random error. In an explanatory IRT framework (De Boeck and Wilson 2004),

---

[2]The amount of bias resulting from a dishonest response can be statistically derived, and the main conclusion is that as a result of the multivariate response structure, a robust estimate of the underlying trait is obtained.

Figure 2
PROBABILITY TREE OF RESPONDING TO AN ORDINAL ITEM UNDER RR

variation in item parameters can be added as well. For example, we could impose that the first threshold varies as a function of certain item characteristics:

$$(7) \qquad \gamma_{k,1} = \boldsymbol{\delta}' \mathbf{W}_k + \psi_k,$$

where $\mathbf{W}_k$ is a vector of item characteristics and ψ_k is a random error. Finally, it is possible to accommodate situations in which a subgroup of respondents does not follow the procedure, thus preventing such respondents from biasing the results.[3] That is, some people might always give a safe, socially desirable response regardless of whether the randomization mechanism tells them to provide a truthful or a forced response. To deal with this, we can specify two distinct latent classes. The first class of respondents will choose the socially safe response no matter what, whereas the second class follows the RR procedure (see Böckenholt and Van der Heijden 2007). Respondents belong to the first class with probability κ and to the second class with probability $1-\kappa$.

In summary, the proposed polytomous item RR model guarantees complete confidentiality to the participants regarding their responses to sensitive questions. It is tailored to the ordinal response scales that are common in marketing and incorporates predictors at the item and person side, which allows for structural analyses. Finally, not everybody necessarily needs to follow the procedure.

Estimation and Model Fit

The likelihood L, under a random sampling of people, can be marginalized over the latent construct. Maximum marginal likelihood methods, in combination with Gauss–Hermite quadrature, can be used, but the order constraints in the thresholds are difficult to embed in the likelihood. Therefore, we opt for a fully Bayesian approach instead, which specifies prior distributions for all model parameters. We specify a slightly informative inverse gamma prior for the variance components and a Beta$(1, 1)$ prior for the parameter κ governing the latent classes. The prior for the discrimination parameters is flat and subject to the constraint that the parameters must be positive (details are available in the Web Appendix A at http://www.marketingpower.com/jmrfeb10).

---

[3]We thank an anonymous reviewer for suggesting this.

The fit of the model and model assumptions can be evaluated with Bayesian residual analysis. Although such analyses have been proposed for binary IRT models, we provide extensions here for the case of polytomous data. The posterior distributions of the residuals provide information about the magnitudes of these residuals, which can be used to make posterior probability statements about the values of the residuals. Here, a Bayesian residual is defined as the difference between the latent response and the expected (latent) response, $r_{ik} = \tilde{Y}_{ik} - \sum_{c=1}^{C} c \times \pi_{ikc}$, which needs to be estimated from the RR data. However, the estimation of the Bayesian residuals is easily integrated in the Markov chain Monte Carlo (MCMC) algorithm for estimating the model parameters (see Web Appendix A at http://www.marketingpower.com/jmrfeb10). In each iteration, Bayesian residuals are computed using the sampled values of the augmented latent response data and the model parameters given the RR data. The computed sequences of Bayesian residuals can be viewed as draws from their marginal posterior distributions. We considered two functions of sums of squared residuals to evaluate the fit of the items and of the people: $Q_k(\mathbf{y}) = \sum_{i=1}^{n} r_{ik}^2$, and $Q_i(\mathbf{y}) = \sum_{k=1}^{K} r_{ik}^2$, respectively. We used both functions as a discrepancy measure in a posterior predictive check:

$$(8) \qquad P[Q_k(\mathbf{y}^{rep}) \geq Q_k(\mathbf{y}) \mid \mathbf{y}]$$

$$= \int I[Q_k(\mathbf{y}^{rep}) \geq Q_k(\mathbf{y})]p(\mathbf{y}^{rep} \mid \mathbf{y})d\mathbf{y}^{rep},$$

where $\mathbf{y}^{rep}$ denotes replicated RR data under the model. The (estimated) extremeness of the realized observed discrepancy is evaluated with replicated data under the model. A relatively high function value corresponds to a poor fit of the corresponding item or person in which the sampling distribution of the replicated data is used to quantify the extremeness of the realized discrepancy.

A test for local independence can be performed by computing the conditional covariance between Bayesian residuals pertaining to two items given the person parameters. That is, let $r_k(\mathbf{y}, \xi)$ denote the vector of residuals for item k given the RR data and the person parameters. Subsequently, the covariance between two vectors of item residuals is denoted as $\sigma_{k,k'}(\mathbf{y}, \xi) = \text{cov}[r_k(\mathbf{y}, \xi), r_{k'}(\mathbf{y}, \xi)]$, which can be computed within each iteration of the MCMC algorithm using the expected a posteriori estimate for the person parameters. The MCMC sequence can be used to estimate $\sigma_{k,k'}(\mathbf{y}, \xi)$, and its corresponding posterior distribution can be used to check whether this estimate is significantly different from zero, which indicates a violation of local independence.

APPLICATION TO ADULT ENTERTAINMENT DESIRES

Adult entertainment is one of the largest and most profitable industries worldwide.[4] Despite its economic and sociocultural impact, surprisingly little is known about individual-level consumption and desires to consume adult entertainment, and there is even less theorizing about the

topic (Cronin and Davenport 2001). Surveys of adult entertainment consumption have mostly used direct questioning to explore isolated, specific topics, such as attitudes toward pornography (Evans-DeCicco and Cowan 2001), searching sexually explicit material on the Internet (Goodson, McCormick, and Evans 2001), or patronage of sex workers (Traen, Eek-Jensen, and Stigum 2005). Adult entertainment represents an ideal setting to test our methodology because it elicits strong self-conscious emotions, such as embarrassment, shame, and guilt. When querying people directly, they might feel the urge to respond in culturally accepted ways, without displaying their deeper sexual desires and activity. Our focus in the survey was on desires because these are intimately linked to deep-seated preferences (Baumeister, Catanese, and Vohs 2001; Schmitt 2003) and influence consumption behavior over extended periods of time (Perugini and Bagozzi 2001, 2004).

Survey of Adult Entertainment Desires

Sample and design. A randomly drawn sample of 1260 members of the CentERdata panel of Tilburg University participated voluntarily in a study about desires for adult entertainment. The panel is nationally representative for people over 16 years of age. Panel members, who live in the Netherlands, received questionnaires electronically, completed them at home on their personal computers, and then returned them.

Panel members were assigned to a condition of a two-group between-subjects experimental design. The direct question group (n = 684) answered questions without any randomization mechanism, and the RR group (n = 576) used a randomized procedure when answering the questions (following the rationale in Figure 1), for which they received a die at home by regular mail, one week before participating in the survey. We illustrate the usefulness of the proposed model by comparing the responses of the two experimental groups on the observed items and the underlying latent desire construct and by relating these to underlying socioeconomic variables of interest. Information was available about participants' gender (male/female), age (<28, 28–45, and >45), and net monthly income (<€700, €700–€1700, >€1700). These variables can test the validity of the method by relating them to adult entertainment desires as main effects and to examine whether and how these characteristics potentially interact with question technique (direct versus RR) to influence reported adult entertainment desires, as we describe subsequently.

Questionnaire and instructions. To lower the threat to participants, we used descriptive terms and natural language for all instructions, items, and other communication with participants, which is further supported by using (nonpersonal) electronic data collection (Tourangeau, Rips, and Rasinski 2000). We developed a multi-item measure of adult entertainment desires, building on the work of Perugini and Bagozzi (2001, 2004) on desires in general and that of Schmitt (2003) on desires for sexual variety more specifically. Perugini and Bagozzi (2004, p. 71) define desires as "a state of mind whereby an agent has a personal motivation to perform an action or achieve a goal." Desires differ from related constructs, such as intentions, because the former are less connected to immediate action

---

[4]See http://internet-filter-review.toptenreviews.com/internet-pornography-statistics.html (last accessed November 2008).

and are framed over longer time horizons, with the feasibility of the action or goal attainment being less relevant. Desires also differ from fantasies because the former have survived some reality check, whereas the latter do not require this. Importantly, there is evidence that all traditional determinants of behavioral intentions and subsequent volitional behavior are mediated by desires (Perugini and Bagozzi 2001).

To cover the domain of adult entertainment desires, we examined the academic literature on the topic (e.g., Cronin and Davenport 2001; Evans-DeCicco and Cowan 2001; Goodson, McCormick, and Evans 2001), Web sites dedicated to the adult entertainment industry (e.g., www.avnonline.com; http://newsstats.sextracker.com; http://internet-filter-review.toptenreviews.com/internet-pornogaphy-statistics.html), and international suppliers of adult entertainment products and services (http://www.beate-uhse.com). On the basis of this examination, we identified 15 desires that span the domain of legal (in the Netherlands) adult entertainment consumption (for further details, see Web Appendix B at http://www.marketingpower.com/jmrfeb10): (1) free Web sites; (2) paid Web sites; (3) pay-per-view television channels in the home; (4) pay-per-view television channels out of the home; (5) magazine and books; (6) videotapes, DVDs, CD-ROMS, and similar media; (7) telephone services; (8) dating sites or telephone services; (9) toys; (10) sex stimulating or prolonging products; (11) erection stimulating products; (12) clothing; (13) clubs; (14) opposite-gender prostitution; and (15) same-gender prostitution.

The items cover, respectively, private consumption of prepackaged products and services (Items 1–6), consumption through interaction with distant others (Items 7–8), instruments in sex activity with others (Items 9–12), and consumption of customized sex services (Items 13–15). Item selection aimed to reflect ordinality in the intensity of adult entertainment desires, such that participants endorsing more difficult items (e.g., Items 13–15) in the set should have a greater probability to endorse easier items (e.g., Items 1–6), similar to the principles of Mokken and Guttman scales. Note that difficulty scaling using a broad spectrum of items has been advocated for sexual behavior scales (Hennessy et al. 2008) to achieve validity (see Grello, Welsh, and Harper 2006; Meston et al. 1998). Although the use of a priori item design is relatively new to the field of marketing, it is consistent with recently developed psychometric models by De Boeck and colleagues (see De Boeck and Wilson 2004). Incorporating item effects introduces dual sources of variation in the observed scores, which facilitates the understanding of how person- and item-related factors contribute to adult entertainment desires, as we show subsequently.

The introduction to the desire measurement instrument in the questionnaires read as follows:

> This questionnaire is about your desires and wishes with respect to paid forms of sex and pornography, and with respect to products and services related to this. Sex is the act of intercourse. Pornography covers all words, images, objects, symbols and so forth that explicitly refer to, describe or depict acts of intercourse. Desires and wishes cover things that you would like to do or achieve, if the circumstances would be right.

For participants in the direct questioning group, this general instruction was followed by the 15 adult entertainment desire items, each with the following sentence stem: "If *no one* would *ever* discover this, then I *would like* to…" (underlining in the original). Respondents assessed these items on five-point response categories: (1) "strongly disagree," (2) "disagree," (3) "neither disagree nor agree," (4) "agree," and (5) "strongly agree."

RR procedure. Participants in the RR group received an additional instruction, directly following the general instruction and before receiving the questions. The instruction on the computer screen read as follows (we used separate screens to parcel the information):

> Because the topic of this survey may be sensitive to you, we want to protect your answers. To this aim, for each of the following questions, you can sometimes give your true answer and sometimes an answer that depends on a die (one dice). We now explain how this works. It is <u>essential</u> that you follow the procedure exactly, even if you have some reservation to provide a truthful answer on a certain question, or when the die provides an outcome that you do not like. Please now take the die that has been sent to you. Procedure:
>
> 1. For each question, you throw the die.
> 2. If the first throw of the die gives a 1, 2, 3, or 4, <u>then please provide your truthful answer to the question.</u>
> 3. If the first throw of the die gives 5 or 6, then throw the die again.
> 4. If the second throw of the die gives a 1, then provide the answer 1 to the question.
> 5. If the second throw of the die gives a 2, then provide the answer 2 to the question.
> 6. If the second throw of the die gives a 3, then provide the answer 3 to the question.
> 7. If the second throw of the die gives a 4, then provide the answer 4 to the question.
> 8. If the second throw of the die gives a 5 or 6, then provide the answer 5 to the question.
>
> By following this procedure carefully, it is impossible for CentERdata to know your real answers to the questions, because <u>only you</u> know the outcome of each throw of the die.
>
> If CentERdata sees a 4 as an answer to a certain question, then this may be because your first throw of the die gave a 1, 2, 3, or 4, and you provided your truthful answer, which was 4, but it could also be that your first throw with the die gave a 5 or 6, and your second throw with the die gave a 4.
>
> This procedure is known as the "Randomized Response" technique, and it is used in research like this. By using this technique, we can make inferences for each question aggregated across respondents (because we correct for the answers due to the die) but not anymore for any individual participant in the survey. CentERdata can therefore never detect what your true answer to any of the following questions is. Therefore, you can, with all your heart, provide your true answer to a question if the die tells you to. Therefore, please follow the instructions that you have just been provide strictly.

Establishing Validity

Relationships with other variables. Improvements in truthful responding as a result of RR techniques can be unequivocally assessed when objective, external data exist to verify the veracity of the survey responses. Yet, such objective, external data are often not available when constructs are intrinsically unobservable, as is common in marketing. Thus, the validity of RR techniques in uncovering consumers' true answers needs to be assessed indirectly by relating the target construct to other variables (Churchill 1979; Roese and Jamieson 1993; for similar arguments in the domain of sexual behavior, see Hennessy et al. 2008). We take this approach here. Specifically, we derive and test predictions about the relationships among gender, age, income, and erotic desires and how these relationships are influenced by social desirability biases. Finding the theoretically derived relationships in the data supports the nomological validity of our methodology.

First, because of its sensitive nature, the overall level of adult entertainment desire should be lower in the direct question group than in the RR group, conveying underreporting in the former. Second, and more specifically, women are expected to report lower levels of adult entertainment desire than men. This is in line with the conclusions from Baumeister, Catanese, and Vohs's (2001) literature review of gender differences in sex drive and with Baumeister and Vohs's (2004) recent social exchange theory of heterosexual interaction, which argues that women are sellers and men buyers of sex and that the former are prone to keep the desired resource in short supply to exert power. Thus, men across countries report overall higher desires for sexual partners than women do (Schmitt 2003), and male students are more likely than female students to report having watched sexually explicit material on the Internet and to have masturbated while watching this (Goodson, McCormick, and Evans 2001). Importantly, however, we expect question technique to interact with gender, such that the increase in reported desire in the RR group will be higher for women than for men. That is, women should not only hold their scarce resource in short supply but are also socialized to communicate the lowered drive, which gives rise to the short supply (Baumeister and Vohs 2004), in turn making reports of their true desire for adult entertainment more sensitive. Thus, more so than men, women should underreport adult entertainment desire in direct questioning compared with RR questioning. However, the final desire of women should still be significantly lower than that of men because of evolutionary and other factors, even under RR.

Third, we expect age to affect desires for adult entertainment in a nonlinear way. That is, desires for adult entertainment should be lowest for the highest age group (>45 years of age), for whom overall sex drive and mating urges are reduced for biological and other reasons, followed by the youngest age group (<28 years of age), who have sufficient opportunities through noncommercial forms of sexual activity and have not yet satiated these. Desires should be highest in the middle age group (28–45 years of age), for whom sexual urges are still strong and satiation lurks in longer but not yet fully stabilized marital and other intimate relationships, as reflected in higher marital infidelity and divorce rates in this age group (Treas and Giesen 2000).

Moreover, we expect question technique to interact with age, such that the increase in desires under RR questioning will be the lowest for the oldest and youngest age groups and the highest for the 28–45 age group. That is, because actual desires for adult entertainment are presumably lowest for the youngest and, in particular, the oldest age group, reporting these lower desires poses little threat to self-esteem. Instead, sensitivity to adult entertainment desires should be much higher for the 28–45 age group because of the implications for strain on marital and other intimate relationships.

Finally, we are not aware of prior research on the role of income in adult entertainment consumption and desires and thus explore its potential influence. Because our items covered free (Item 1), low-priced (e.g., Items 3, 5, and 6), and higher-priced forms of adult entertainment (e.g., Items 13, 14, and 15), the income–desire relationship is another noteworthy topic for exploration.

Effects of instruction. Improvements in truthful responding due to RR techniques could be caused merely by the different introduction section compared with the direct questioning condition (Tourangeau, Rips, and Rasinski 2000). Indeed, the detailed introduction could act a signal that privacy concerns and responses to the survey are taken seriously, which by itself might induce respondents to answer more truthfully than respondents in the direct questioning condition. In this case, improvements in truthful responding in the RR group would be partly due to other, namely instructional, factors, which would compromise construct validity.

To examine this possibility, we randomly split the direct questioning group into two subgroups, one (n = 336) receiving the standard instructions before the specific information about the survey topic and the other (n = 347) receiving a more extended version of the instructions. We described the standard instructions (35 words) previously. The extended instructions (100 words) also emphasized that respondents' participation in the study was essential, that relevant new insights might be gained because of the specific information provided, and that strict confidentiality in handling the responses to the questions was assured.

Analyses of variance revealed no differences between the two instruction groups on the desire for adult entertainment construct (mean latent desire = .022 versus mean latent desire = −.023, $p > .50$) nor any significant interaction effects between the two instruction groups and the sociodemographic variables on the construct (all $Fs < 1.00$). We conclude that priming truthfulness and confidentiality by providing specific instructions to respondents did not influence the target construct in this survey, and therefore we decided to analyze the data across the two instruction groups.

RESULTS

Effects of Question Technique and Socioeconomic Characteristics

Model 1 estimates the main effect of question technique on adult entertainment desires. Models 2–4 estimate the interaction between question technique and, respectively, gender, age, and income, while controlling for their main effects on adult entertainment desires. In all models, the estimated Bayesian residuals did not show a significant

percentage of extreme residuals, and the estimated item discrepancy measure did not detect any item misfits. Less than 5% of the respondents were detected as misfits, meaning that their response patterns can be considered extreme under the model. We detected no significant violations of local independence. Taken together, this implies that the scale is unidimensional and that the model fits well.

For Model 1, the latent scores of the direct questioning group and the RR group are simultaneously derived through the graded response IRT model. The difference is that for the RR group, there is an additional randomization process that influences the observed scores, but this is incorporated into the MCMC algorithm. Because participants were randomly assigned to conditions of the experimental design, differences in group means can only be due to the RR technique. Regressing latent desire ξ on a dummy variable (RR), which is 1 for participants in the RR group and 0 otherwise, reveals that in the RR condition, participants report significantly higher desires for adult entertainment than under direct questioning (posterior mean $\beta_{RR} = .507$, zero not included in the 95% credible interval; see Model 1 in Table 1). In addition, the latent class structure in the model indicates that the number of people who do not follow the instructions is fairly small (posterior mean $\kappa = 10.3\%$). Although this latent class size is fairly small, not accounting for it would reduce the estimate of the difference in mean latent desire by .152 to an estimate of .355, which is large. These findings are in line with other research that has related univariate and multivariate binary direct questioning data with RR data showing the validity increases under such designs (Fox and Tracy 1986; Fox 2005; Lamb and Stem 1978; Lensvelt-Mulders et al. 2005; Locander, Sudman, and Bradburn 1976).

Table 1
INFLUENCE OF QUESTION TECHNIQUE AND SOCIOECONOMIC CHARACTERISTICS ON ADULT ENTERTAINMENT DESIRE

Predictors	Influence on Adult Entertainment Desires			
	Model 1	Model 2	Model 3	Model 4
Question technique (RR = 1) (β_{RR})	**.507**	**.493**	**.545**	**.497**
I (age < 28) ($\beta_{(age < 28)}$)[a]		**−.263**	**−.455**	**−.308**
I (age > 45) ($\beta_{(age > 45)}$)		**−.567**	**−.603**	**−.560**
Gender (female = 1) (β_{Gender})		**−.740**	**−.646**	**−.631**
I (income < 700) ($\beta_{(income < 700)}$)		**−.208**	**−.181**	**−.346**
I (income > 1700) ($\beta_{(income > 1700)}$)		−.060	−.030	−.045
Interactions				
RR × gender ($\beta_{RR,Gender}$)		**.199**		
RR × I (age < 28) ($\beta_{RR,(age < 28)}$)			**.484**	
RR × I (age > 45) ($\beta_{RR,(age > 45)}$)			.102	
RR × I (income < 700) ($\beta_{RR,(income < 700)}$)				**.446**
RR × I (income > 1700) ($\beta_{RR,(income > 1700)}$)				.039
Latent class probability (κ) (%)	10.3	10.3	10.0	9.7

[a]The I(·) function is an indicator function that equals 1 if the logical expression in it is true and 0 if otherwise. Thus, I (age < 28) would be 1 for age < 28 and 0 for age ≥ 28.

Notes: Bold estimates do not contain 0 in their 95% credible interval.

We estimated Models 2–4 to investigate potential biasing effects on the relationships between socioeconomics and desire. In the models, we included main effects of question technique and sociodemographics, and we added the interactions between question technique (dummy RR) and a particular socioeconomic variable, one at a time. In all models, the estimates for the latent class parameter are similar to Model 1.

Model 2 shows that in support of the predictions, women have a clearly lower desire for adult entertainment than men (posterior mean $\beta_{gender} = -.740$, zero not included in the 95% credible interval). However, and also in support of the predictions, the gender desire gap visibly narrows under the RR condition (posterior mean $\beta_{RR,gender} = .199$, zero not included in the 95% credible interval). Tracy and Fox (1981), who investigated self-reported police arrests, also find that women are more likely to underreport. So, perhaps the tendency might transcend specific domains and generalize to norm-deviating behavior more generally. Yet, even under RR, women have less of a desire for adult entertainment, as we hypothesized.

Model 3 documents the age effect, with the middle age group (28–45 years) as the baseline. The analysis shows that this group indeed has the highest desire, in support of our hypothesis. More specifically, the younger and, in particular, the older age groups score significantly lower than the middle age group (posterior means are, respectively, $\beta_{(age < 28)} = -.455$ and $\beta_{(age > 45)} = -.603$, zero not in the 95% credible interval). The interaction with question technique indicates that the youngest participants have the highest increase in reported desire ($\beta_{RR,(age < 28)} = .484$), which is unexpected. The gap in adult entertainment desires between the youngest and the middle age groups disappears completely under RR. This reflects that younger consumers have an equally strong adult entertainment desire but are more susceptible to normative pressures to underreport this under direct questioning, perhaps because desires for commercial forms of sexual activity are perceived to indicate a failure to sustain healthy, noncommercial, sexually satisfying relationships.

Finally, Model 4 reveals a main effect of income and an interaction effect of question technique and income on adult entertainment desires. Specifically, compared with the other income groups, lower-income participants report significantly lower desires for adult entertainment ($\beta_{(income < 700)} = -.346$, zero not in the 95% credible interval), but the strength of their desires increases significantly under RR ($\beta_{(RR,income < 700)} = .446$, zero not in the 95% credible interval). This finding may reflect the normative pressures perceived by lower-income groups to adhere to the social norm against consumption and desires for adult entertainment and how the RR technique alleviates these strains. Although the finding is new, it is consistent with Tracy and Fox's (1981) finding that low-income consumers have greater tendencies to underreport illegal activities.

Item-Level Desires for Adult Entertainment

We intentionally selected items to reflect different content and different intensities (or "difficulties") of desire for adult entertainment (Hennessy et al. 2008). That is, all items tap a libidinistic and utilitarian appetite, but it is much

Table 2

OPERATING CHARACTERISTICS OF ADULT ENTERTAINMENT DESIRE ITEMS

	Item Characteristics				
Adult Entertainment Desire Items	Discrimination	Threshold 1	Threshold 2	Threshold 3	Threshold 4
Item 1: free pornographic Web sites	1.301	.059	.545	1.077	1.680
Item 2: paid pornographic Web sites	1.985	1.013	1.465	2.022	2.570
Item 3: pay-per-view television channels in the home	1.443	.886	1.341	1.908	2.451
Item 4: pay-per-view television channels out of the home	2.002	1.079	1.570	2.115	2.695
Item 5: pornographic magazines or books	1.991	.328	.803	1.385	1.927
Item 6: pornographic videos, DVDs, and so forth	1.996	.296	.656	1.282	1.894
Item 7: erotic telephone services	1.472	1.335	1.815	2.364	3.003
Item 8: sex dating sites or telephone services	1.212	1.612	2.152	2.652	3.332
Item 9: sex toys	.905	.365	.869	1.428	2.007
Item 10: sex stimulants and prolongations	1.218	1.001	1.486	1.921	2.581
Item 11: erection stimulants	1.604	.905	1.357	1.968	2.583
Item 12: erotic clothing	1.083	.643	1.167	1.740	2.357
Item 13: sex clubs	1.173	1.061	1.568	2.128	2.671
Item 14: opposite-gender prostitution	1.813	1.147	1.635	2.153	2.800
Item 15: same-gender prostitution	.768	2.881	3.444	3.947	4.555

Notes: Item number indicates the position of the item in the survey (i.e., Item 1 means that this item was the first item administered).

"easier" for a person to browse free Web sites in the privacy of his or her own home (Item 1) than to pay a visit to a specialized sex club (Item 13). In theory, only participants who are very high on desire would yearn for more difficult forms of commercial sex. Table 2 presents the posterior means of the discrimination and threshold parameters for all adult entertainment desire items for participants in the RR condition. Items 2, 4, 5, 6, and 14 best reflect the desire construct, as shown by their high discrimination parameters. The threshold parameters are particularly noteworthy because they reflect the "difficulty" of passing a certain response option. For example, if a participant has a desire value higher than the first threshold, it means that the participant is likely to have answered 2 or higher on the response scale. Item 1 (free Web sites) has the lowest and Item 15 (same-gender prostitution) has the highest first threshold value. All other items fall somewhere in between the first and the last items, lending support for our choice of items.

To know the percentage of consumers choosing a 4 or 5 response on a particular item, simulations must be used. We simulated data based on our model to establish how many people would indicate agreement, or strong agreement, with the statement that they would like to visit free Web sites if nobody would find out. The percentages under direct questioning and RR appear in Table 3 and are based on a simulation using 1,000,000 respondents to examine population proportions.[5] We obtain the direct questioning percentages by shifting the desire trait with an amount equal to the mean difference between the two experimental groups and using the same item parameters. Under RR, approximately 25% of the participants would like to display this type

of behavior. This shows a substantial difference compared with the direct questioning group. For some of the other items, the differences in agreement are also large. Note that the largest differences between the direct questioning and the RR groups are for desire items of the highest difficulty.

Other noteworthy findings are revealed when we consider the item-level predictors, which is a completely different perspective on the data. Similar to De Boeck and Wilson (2004), we used an a priori item design with groups

Table 3

AGREEMENT WITH THE DESIRE ITEMS FOR THE DIRECT QUESTION AND RR GROUPS

	Percentage Agreement	
Adult Entertainment Desires: "I would like to consume…"	Direct Question Group (%)	RR Group (%)
Item: free pornographic Web sites	14.9	24.7
Item: paid pornographic Web sites	10.1	18.7
Item: pay-per-view television channels in the home	7.6	14.2
Item: pay-per-view television channels out of the home	9.3	17.5
Item: pornographic magazines or books	15.9	27.5
Item: pornographic videos, DVDs, and so forth	16.5	28.3
Item: erotic telephone services	4.7	9.4
Item: sex dating sites or telephone services	1.8	3.9
Item: sex toys	9.0	14.9
Item: sex stimulants and prolongations	5.5	10.5
Item: erection stimulants	7.9	14.9
Item: erotic clothing	6.7	12.1
Item: sex clubs	4.6	8.8
Item: opposite-gender prostitution	7.5	14.3
Item: same-gender prostitution	.03	.08

Notes: Agreement is response 4 ("agree") or 5 ("strongly agree").

[5]Interest is in population proportions, and with a small sample, there can be substantial variation. In addition, the threshold values and the latent desire trait are measured on the same scale, and a large first threshold for an item implies that very few people would display the desire as described in the item. Thus, a small sample will not reliably indicate the population proportions.

of items. In other words, we intentionally constructed the scale on the basis of item properties, and they are an integral part of our "test blueprint." This enables us to explore whether certain item properties are related to the threshold values, which we do by entering one covariate at a time into the first threshold equation.[6] Item groups were consumption of prepackaged products and services (Items 1–6), consumption through interaction with distant others (Items 7–8), instruments in sex activity with others (Items 9–12), and consumption of customized sex services (Items 13–15). Of these four groups, the effect of prepackaged products and services and the effect of customized service were significantly related to the first threshold. The posterior mean of $\delta_{prepackaged}$ is $-.71$ (SE $= .36$), while the posterior mean δ_{custom} is 1.03 (SE $= .40$). Not unexpectedly, this reveals that participants find it difficult to give a high score to items that involve explicit sexual behavior with other people, while it is much easier for them to simply buy products and services to enhance sexual activity.

Dedicated SDR Scales

Although there are problems with dedicated SDR scales because they tap both trait content and response style and though measuring style but not content is their aim (Paulhus 2002; Smith and Ellingson 2002), we nonetheless also performed an analysis with these scales because they are often used in marketing as covariates in direct question surveys (e.g., Bearden and Netemeyer 1999; Mick 1996). We assessed social desirability in a separate wave (three weeks interspersed) among the direct question group using the impression management (IM) scale of Paulhus's balanced inventory of desirable responding (see also Mick 1996). Prior research indicates that IM might affect SDR in sexual matters (Meston et al. 1998). The scale consists of 20 five-point "disagree/agree" items, but because of the length of the instrument, we administered only a subset of 10 items to measure social desirability (we omitted potentially offensive and/or inappropriate items while retaining the balanced structure of the scale; 5 positively and 5 negatively worded items). The reliability of our IM measure is .70, which is acceptable. We used the average IM score (M $= 3.20$, SD $= .62$) in the analyses, after reversing negatively formulated items such that higher scores indicate higher levels of IM.

We examined the relationship between the desire construct and sociodemographic variables using regression analysis. To the extent that the relationships between desire and these other variables change significantly after accounting for the social desirability measures, these relationships are biased (Podsakoff et al. 2003; Tourangeau, Rips, and Rasinski 2000). To examine this possibility, in Step 1 of the analysis, we regressed the scores on the desire construct on the five dummy variables representing the sociodemographic variables. Next, in Step 2, we added the score on the IM scale (see Table 4). In Step 2, IM was significantly related to the desire construct (estimate $-.275$, t $= -4.66$, $p < .001$), indicating that expressed desires for adult entertainment are indeed reduced when IM is higher. Crucially,

---

[6]Although it would be preferable to include all item predictors simultaneously, the number of observations is too small for that.

Table 4
INFLUENCE OF SDR SCALE AND SOCIOECONOMIC CHARACTERISTICS ON ADULT ENTERTAINMENT DESIRE IN DIRECT QUESTION GROUP

Predictors	Influence on Adult Entertainment Desires		
	Model 1	Model 2	Model 3
Intercept (β_0)	1.185 (.133)	1.983 (.198)	2.377 (.674)
I (age < 28) ($\beta_{(age\ <\ 28)}$)[a]	$-.391$ (.130)	$-.460$ (.127)	$-.598$ (.681)
I (age > 45)($\beta_{(age\ >\ 45)}$)	$-.505$ (.072)	$-.460$ (.071)	$-.425$ (.370)
Gender (female = 1) (β_{Gender})	$-.547$ (.072)	$-.511$ (.071)	$-.791$ (.375)
I (income < 700) ($\beta_{(income\ <\ 700)}$)	$-.188$ (.104)	$-.165$ (.102)	$-.505$ (.619)
I (income > 1700) ($\beta_{(income\ >\ 1700)}$)	$-.043$ (.079)	$-.037$ (.077)	.026 (.414)
IM (β_{IM})		$-.275$ (.052)	$-.398$ (.210)
Interactions			
IM × gender ($\beta_{RR,Gender}$)			.087 (.229)
IM × I (age < 28) ($\beta_{RR,(age\ <\ 28)}$)			.050 (.114)
IM × I (age > 45) ($\beta_{RR,(age\ >\ 45)}$)			$-.011$ (.116)
IM × I (income < 700) ($\beta_{RR,(income\ <\ 700)}$)			.101 (.186)
IM × I (income > 1700) ($\beta_{RR,(income\ >\ 1700)}$)			$-.019$ (.128)

[a]The I(·) function is an indicator function that equals 1 if the logical expression in it is true and 0 if otherwise. Thus, I (age < 28) would be 1 for age < 28 and 0 for age $\geq$ 28.

Notes: n = 684. Parameter estimates with standard error.

however, none of the parameter estimates of the sociodemographic variables changed significantly between Step 1 and Step 2. That is, all significant t-values remained significant, and all insignificant t-values remained insignificant; there were no sign changes; and the largest absolute difference in t-value was .6.

Finally, in Step 3, we included all five (one for gender, two for age, and two for income) interactions between the IM scale and the sociodemographics. None of the interactions were significant (all *p*s > .11), regardless of whether they were entered separately or jointly as a group. This shows that though social desirability tendencies, as assessed by the IM scale, were significantly associated with the desire construct, they did not significantly change its relationships with important predictors, whereas questioning technique did.

DISCUSSION

Surveys with self-report measures obtained from consumers are routinely employed in marketing research to inform decision makers about important managerial issues. In view of the importance of surveys, it is no surprise that issues in data collection and survey validity have been of long-standing interest to marketers. Among the most pervasive response styles is SDR (Paulhus 2002), especially, but not only, for sensitive questions. We introduced a new item RR model into the field of marketing and applied it to the adult entertainment category. The model allows for (1) polytomous measures, (2) individual-level inferences, (3) structural analyses of potential individual-level determinants and consequences of these constructs, (4) item predictors, (5) a group of respondents who always give socially

desirable answers (i.e., they do not follow the RR instructions), and (6) easy checking of model fit. Our empirical application demonstrated not only that respondents found it difficult to openly admit having desires for adult entertainment and were likely to give untruthful answers if queried directly but also that truthfulness systematically improved under randomized responding and in directions predicted by substantive theory.

When and How to Use Item RR Models

Implementation of the item RR technique can be costly. Our item RR method requires dice, and the required sample size is relatively large (in the hundreds). In addition, careful attention should be paid to issues of comprehension, and clear sets of instructions are important for participants to understand their task. This may make the technique less suitable for less educated or very young consumers or for special populations, such as social welfare offenders (Lensvelt-Mulders et al. 2005). The current application was on the general population over the age of 18, which alleviates these concerns. Finally, the analyses cannot yet be performed using standard software packages.

There are some ways to reduce the costs of RR. In our research, we sent dice to respondents. Nonetheless, there might be cheaper alternatives, such as the use of virtual dice on the Internet or sending preannouncements to respondents in which they are told to obtain a dice (either they already have dice or they can obtain them from a neighbor or friend). We conducted an informal study in three regions (United Kingdom: N = 277; Singapore: N = 264; and Poland: N = 282 [all general population]) to explore how many people are able to obtain dice easily. The market research firm Survey Sampling International conducted the fieldwork. The percentages in all countries were higher than 80%, indicating that cost savings on the dice are possible, at least in these and similar countries. In addition, other randomization devices, such as coin tosses, spinners, or cards, can easily replace dice.

We believe that RR techniques should be seriously considered when it is important to know the exact levels of the latent construct of interest and the relationships between the construct and other variables, the construct's determinants or consequences, and when the implications of deviations from true levels and relationships would be serious. Dedicated social desirability scales cannot accomplish this. Sensitive settings are common, especially in the field of transformative consumer behavior, such as tobacco, alcohol, and drug consumption; gambling; and financial and medical decision making (see Mick 2005). To gauge the potential social sensitivity of surveys, and thus the need to use RR techniques, expert panels or small-scale pilot surveys can be applied, in which the potential shamefulness or embarrassment associated with questions about the topic are determined.

Ethical Considerations

Honesty is paramount to warrant sustained trust in marketing research by the respondents and society at large. The proposed new item RR model provides increased insight into consumers' true responses to sensitive questions. Truthfulness in data collection and confidentiality in data utilization are two central ethical considerations.

Truthfulness. Our survey avoided deception, which otherwise could have been a source of respondent mistrust. Survey respondents can be misled, for example, by suggesting to them that a procedure or measure cannot do what it actually can or by suggesting that it can do what it actually cannot. Randomized response techniques, including our procedure, cannot infer the true answers of individual respondents to each sensitive question, and the introduction in the survey explained this. This privacy protection stimulates respondents to give honest answers to potentially embarrassing questions. With classic RR techniques, only the aggregate responses of the sample as a whole to the sensitive questions can be inferred, not the disaggregate responses of individual respondents to the specific sensitive questions. By applying our item RR model, researchers can now infer the disaggregate scores of respondents on the latent construct of interest, but respondents' true answers to the each specific sensitive question can still not be inferred. Thus, anonymity is preserved at the item level, and respondents are truthfully informed about this, which is important for ethical, methodological, and disciplinary reasons (Kimmel and Smith 2001).

This conforms to the codes of conduct of professional marketing and survey research organizations. For example, the American Association for Public Opinion Research, the professional organization of public opinion and survey research professionals in the United States, states in its code of professional ethics and practices (http://www.aapor.org/aaporcodeofethics) that "1. We shall avoid practices or methods that may harm, humiliate, or seriously mislead survey respondents, and 2. We shall respect respondents' concerns about their privacy." In a similar vein, the European Society for Opinion and Marketing Research, states the following in its code of conduct (http://www.esomar.org/index.php/professional-standards.html): "(a) Market researchers shall not abuse the trust of respondents or exploit their lack of experience or knowledge. (b) Researchers shall not make false statements about their skills, experiences or activities, or about those of their organization." Respondents in RR studies using our model should be clearly and fully informed about which information can and cannot be obtained from their responses to the questions.

However, the use of the bogus pipeline, which has been advocated as an alternative technique to alleviate social sensitivity (Aguinis, Pierce, and Quigley 1993; Derakshan and Eysenck 2005; Roese and Jamieson 1993; Tourangeau, Rips, and Rasinsky 2000) is deceptive and violates the codes of conduct of professional organizations. By using these techniques, respondents are misled to believe that researchers and their organizations can do what they actually cannot. In a typical bogus pipeline study, respondents are connected to a fake lie detector and are informed that it can sense dishonest responses, which it actually cannot. A meta-analysis across 31 studies in various domains has shown that the technique is an effective means of lowering SDR (Roese and Jamieson 1993). Yet, it asks researchers to deceive respondents to prevent the respondents from lying to the researchers, which is ironic. Such a

breach of trust seems to violate professional codes of conduct, endangers the reputation of marketing research, and can be harmful to the respondent (Frankel 1976). Whether deception is acceptable in academic marketing research under controlled, laboratory conditions, with full disclosure of information after completion, informed consent, and other safeguards, is a different issue (Kimmel and Smith 2001). Some reports indicate that students—typical potential participants of academic marketing research—perceive the bogus pipeline as a useful and acceptable technique if the topic under study is important (Aguinis and Henle 2001). Even then, the cost of using fake lie detectors in large-scale consumer and marketing research is a practical drawback of the technique (Roese and Jamieson 1993).

Confidentiality. No information about individual identifiable respondents in our survey was disclosed to third parties, and respondents were informed about this. That is, the survey data were used only for model building and theory testing about the levels of constructs for consumer groups and relationships between them, and these data were reported—not the data of identifiable respondents. Confidentiality stimulates respondents to give honest answers to sensitive questions. The specific survey and research organization (CentERdata; see www.centerdata.nl) complies with the codes of conduct for survey marketing research and international privacy laws and regulations. Clients only receive aggregate reports and data sets without identity information (e.g., name, address, location) of individual respondents, and all researchers sign a nondisclosure form. This conforms to the code of standards and ethics for survey research of the Council of American Survey Research organizations (www.casro.org) and the Marketing Research Association, including the stipulation that researchers "[w]ill not reveal any information that could be used to identify clients without their written authorization. Proper authorization from a client should be in written format prior to or during the data collection process. This authorization must include that the information will only be used for research purposes" (see http://www.mranet.org/about/codes.cfm [point 30]). This also ensures that the survey information cannot be used for nonresearch purposes, such as direct marketing, nonconsent list generation, credit rating, push polling, fund raising, or other marketing or political activities. In particular, in marketing research on sensitive topics, such as that of the current study, strict confidentiality is paramount to ensure respondents' current trust and avoid any potential harm.

Further Research

Given the complexity of SDR, there is still much to learn. We studied only a single domain; differences between direct questioning and item RR in other domains may differ. In addition, investigating the feasibility of electronic randomization devices in Internet data collection is important. Such electronic versions of randomization devices might evoke suspicion among participants because it is well known that almost anything can be stored digitally. If such electronic versions of randomization devices could be made acceptable to participants, they could strongly contribute to the widescale use of RR techniques in the rapidly expanding Internet and Web panels in marketing research. Trust and effectiveness could also vary cross-culturally.

Although much remains unknown, we hope that our model and empirical application stimulate marketing to pay more attention to social sensitivity in research and to deal with it appropriately.

REFERENCES

Aguinis, Herman and Christine A. Henle (2001), "Empirical Assessment of the Ethics of the Bogus Pipeline," *Journal of Applied Social Psychology*, 31 (2), 352–75.

———, Charles A. Pierce, and Brian M. Quigley (1993), "Conditions Under Which a Bogus Pipeline Procedure Enhances the Validity of Self-Reported Cigarette Smoking: A Meta-Analytic Review," *Journal of Applied Social Psychology*, 23 (5), 352–73.

Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103 (3), 411–23.

Baumeister, Roy F., Kathleen R. Catanese, and Kathleen D. Vohs (2001), "Is There a Gender Difference in Strength of Sex Drive? Theoretical Views, Conceptual Distinctions, and a Review of Relevant Evidence," *Personality and Social Psychology Review*, 5 (3), 242–73.

——— and Kathleen D. Vohs (2004), "Sexual Economics: Sex as Female Resource for Social Exchange in Heterosexual Interactions," *Personality and Social Psychology Review*, 8 (4), 339–63.

Bearden, William O. and Richard G. Netemeyer (1999), *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research*, 2d ed. Newbury Park, CA: Sage Publications.

Böckenholt, Ulf and Peter G.M. Van der Heijden (2007), "Item Randomized–Response Models for Measuring Noncompliance: Risk–Return Perceptions, Social Influences, and Self-Protective Responses," *Psychometrika*, 72 (June), 245–62.

Bruner, Gordon C. and Paul J. Hensel (1992), *Marketing Scales Handbook: A Compilation of Multi-Item Measures*. Chicago: American Marketing Association.

Churchill, Gilbert A., Jr. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research*, 16 (February), 64–73.

Clark, Stephen J. and Robert A. Desharnais (1998), "Honest Answers to Embarrassing Questions: Detecting Cheating in the Randomized Response Model," *Psychological Methods*, 3 (June), 160–68.

Cronin, Blaise and Elisabeth Davenport (2001), "E-Rogenous Zones: Positioning Pornography in the Digital Economy," *The Information Society*, 17 (1), 33–46.

De Boeck, Paul and Mark Wilson (2004), *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York: Springer-Verlag.

De Jong, Martijn G., Jan-Benedict E.M. Steenkamp, and Jean-Paul Fox (2007), "Relaxing Measurement Invariance in Cross-National Consumer Research Using a Hierarchical IRT Model," *Journal of Consumer Research*, 34 (August), 260–79.

———, ———, ———, and Hans Baumgartner (2008), "Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation," *Journal of Marketing Research*, 45 (February), 104–115.

Derakshan, Nazanin and Michael W. Eysenck (2005), "When the Bogus-Pipeline Interferes with Self-Deceptive Strategies: Effects on State Anxiety in Repressors," *Cognition and Emotion*, 19 (1), 83–100.

Evans-DeCicco, Jennee A. and Gloria Cowan (2001), "Attitudes Toward Pornography and the Characteristics Attributed to Pornography Actors," *Sex Roles*, 44 (5–6), 351–61.

Fischer, Robert J. (1993), "Social Desirability Bias and the Validity of Indirect Questioning," *Journal of Consumer Research*, 20 (September), 303–315.

Fox, James A. and Paul E. Tracy (1986), *Randomized Response: A Method for Sensitive Surveys*. Beverly Hills, CA: Sage Publications.

Fox, Jean-Paul (2005), "Randomized Item Response Theory Models," *Journal of Educational and Behavioral Statistics*, 30 (2), 189–212.

Frankel, Lester R. (1976), "Statistics and People: Statisticians Responsibility," *Journal of the American Statistical Association*, 71 (353), 9–16.

Goodson, Patricia, Deborah McCormick, and Alexandra Evans (2001), "Searching for Sexually Explicit Materials on the Internet: An Exploratory Study of College Students' Behavior and Attitudes," *Archives of Sexual Behavior*, 30 (2), 101–118.

Greenberg, Bernard G., Abdel-Latif A. Abul-Ela, Walt R. Simmons, and Daniel G. Horvitz (1969), "Unrelated Question Randomized Response Model: Theoretical Framework," *Journal of the American Statistical Association*, 64 (326), 520–39.

Grello, Catherine M., Deborah P. Welsh, and Melinda S. Harper (2006), "No Strings Attached: The Nature of Casual Sex in College Students," *Journal of Sex Research*, 43 (3), 255–67.

Hennessy, Michael, Amy Bleakley, Martin Fishbein, and Amy Jordan (2008), "Validating an Index of Adolescent Sexual Behavior Using Psychosocial Theory and Social Trait Correlates," *AIDS and Behavior*, 12 (2), 321–31.

Hill, Peter C., Charles A. Dill, and Ernest C. Davenport (1988), "A Reexamination of the Bogus Pipeline," *Educational and Psychological Measurement*, 48 (3), 587–601.

Jones, Edward E. and Harold Sigall (1971), "The Bogus Pipeline: A New Paradigm for Measuring Affect and Attitude," *Psychological Bulletin*, 76 (5), 349–64.

Kimmel, Allan J. and N. Craig Smith (2001), "Deception in Marketing Research: Ethical, Methodological, and Disciplinary Implications," *Psychology & Marketing*, 18 (7), 663–89.

Lamb, Charles W. and Donald E. Stem (1978), "An Empirical Validation of the Randomized Response Technique," *Journal of Marketing Research*, 15 (November), 616–21.

Lensvelt-Mulders, Gerty J.L.M., Joop J. Hox, Peter G.M. Van der Heijden, and Cora J.M. Maas (2005), "Meta-Analysis of Randomized Response Research: 35 Years of Validation," *Sociological Methods and Research*, 33 (3), 319–48.

Locander, William, Seymour Sudman, and Norman Bradburn (1976), "An Investigation of Interview Method, Threat, and Response Distortion," *Journal of the American Statistical Association*, 71 (June), 269–75.

Lord, Frederic M. and Melvin R. Novick (1968), *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Meston, Cindy M., Julia R. Heiman, Paul D. Trapnell, and Delroy L. Paulhus (1998), "Socially Desirable Responding and Sexuality Self-Reports," *Journal of Sex Research*, 35 (2), 148–57.

Mick, David G. (1996), "Are Studies of Dark Side Variables Confounded by Socially Desirable Responding? The Case of Materialism," *Journal of Consumer Research*, 23 (September), 106–119.

——— (2005), "Meaning and Mattering Through Transformative Consumer Research," presidential address at the Conference of the Association for Consumer Research, San Antonio, TX, (September 28–30).

Paulhus, Delroy L. (2002), "Socially Desirable Responding: The Evolution of a Construct," in *The Role of Constructs in Psychological and Educational Measurement*, H.I. Braun, D.N. Jackson, and D.E. Wiley, eds. Mahwah, NJ: Lawrence Erlbaum Associates, 49–69.

Perugini, Marco and Richard Bagozzi (2001), "The Role of Desires and Anticipated Emotions in Goal-Directed Behaviours: Broadening and Deepening the Theory of Planned Behaviour," *British Journal of Social Psychology*, 40 (1), 79–98.

——— and ——— (2004), "The Distinction Between Desires and Intentions," *European Journal of Social Psychology*, 34 (1), 69–84.

Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff (2003), "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology*, 88 (October), 879–903.

Reinmuth, James E. and Michael D. Geurts (1975), "The Collection of Sensitive Information Using a Two-Stage, Randomized Response Model," *Journal of Marketing Research*, 12 (November), 402–407.

Roese, Neal J. and David W. Jamieson (1993), "Twenty Years of Bogus Pipeline Research: A Critical Review and Meta-Analysis," *Psychological Bulletin*, 114 (2), 363–75.

Samejima, Fumiko (1969), "Estimation of Latent Ability Using a Response Pattern of Graded Scores," *Psychometrika Monograph Supplement*, 17, 1–100.

Schmitt, David P. (2003), "Universal Sex Differences in the Desire for Sexual Variety: Tests from 52 Nations, 6 Continents, and 13 Islands," *Journal of Personality and Social Psychology*, 85 (1), 83–104.

Smith, D. Brent and Jill E. Ellingson (2002), "Substance Versus Style: A New Look at Social Desirability in Motivating Contexts," *Journal of Applied Psychology*, 87 (2), 211–19.

Steenkamp, Jan-Benedict E.M., Martijn G. de Jong, and Hans Baumgartner (2010), "Socially Desirable Response Tendencies in Survey Research," *Journal of Marketing Research*, 47 (April), forthcoming.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski (2000), *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.

——— and Tom W. Smith (1996), "Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context," *Public Opinion Quarterly*, 60 (2), 275–304.

Tracy, Paul E. and James E. Fox (1981), "The Validity of Randomized Response for Sensitive Measurements," *American Sociological Review*, 46 (April), 187–200.

Traen, Bente, Line M. Eek-Jensen, and Hein Stigum (2005), "Sex Customers in Norway 2002," *Electronic Journal of Human Sexuality*, 8 (September 26), 1–18.

Treas, Judith and Deirdre Giesen (2000), "Sexual Infidelity Among Married and Cohabiting Americans," *Journal of Marriage and the Family*, 62 (1), 48–60.

Warner, Stanley L. (1965), "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical Association*, 60 (309), 63–69.