

# State-Dependence Effects in Surveys

Martijn G. de Jong

Erasmus School of Economics, Erasmus University, 3000 DR Rotterdam, The Netherlands,  
mgdejong@ese.eur.nl

Donald R. Lehmann, Oded Netzer

Columbia Business School, Columbia University, New York, New York 10027  
{drl2@columbia.edu, on2110@columbia.edu}

In recent years academic research has focused on understanding and modeling the survey response process. This paper examines an understudied systematic response tendency in surveys: the extent to which observed responses are subject to state dependence, i.e., response carryover from one item to another independent of specific item content. We develop a statistical model that simultaneously accounts for state dependence, item content, and scale usage heterogeneity. The paper explores how state dependence varies by response category, item characteristics, item sequence, respondent characteristics, and whether it becomes stronger as the survey progresses. Two empirical applications provide evidence of substantial and significant state dependence. We find that the degree of state dependence depends on item characteristics and item sequence, and it varies across individuals and countries. The article demonstrates that ignoring state dependence may affect reliability and predictive validity, and it provides recommendations for survey researchers.

*Key words:* state dependence; response process; surveys; scale usage; response styles; survey response models; item response theory; validity; reliability

*History:* Received: August 29, 2010; accepted: March 21, 2012; Eric Bradlow and then Preyas Desai served as the editor-in-chief and Wayne DeSarbo served as associate editor for this article. Published online in *Articles in Advance* July 12, 2012.

## Introduction

In recent years scholars have focused on understanding and modeling the survey response process. Whereas traditionally it was assumed that an observed score on an item could be decomposed into a true score (based on item content) and a random error, the response process is considerably more complex. In particular, several systematic tendencies can occur during the response process (Podsakoff et al. 2003). The most commonly studied response tendency is scale usage (also called response style; see, e.g., Baumgartner and Steenkamp 2001, Greenleaf 1992, Rossi et al. 2001, van Rosmalen et al. 2010). Common examples are acquiescence, extreme responding, and avoidance of extreme responding.

A separate line of research in marketing and economics has examined state-dependence or inertia effects in choice (e.g., Chintagunta 1998, Heckman 1981, Seetharaman et al. 1999). This research has found that choices are serially correlated over time because of positive state dependence beyond the effect of individual preferences. That is, purchasing a product in the last period increases the likelihood that the customer will buy the same product again in the current time period.

In the context of surveys, we define state dependence as the tendency to stick with the response given

to the previous item, regardless of the content of the item and the individual's scale usage. An intriguing question is whether a systematic response tendency in the form of state dependence is present in survey responses. The notion of state dependence is related to previously studied concepts of "response carryover" (e.g., Bickart 1993, Tourangeau et al. 2000), "nondifferentiation" (Krosnick 1991), "patterned" responses and "proximity error" (e.g., Guilford 1954, Kane 1969, Yamamoto 1995, Zedeck and Blood 1974), as well as to the research on common stimulus material as in the testlet theory (Bradlow et al. 1999; Wainer et al. 2007; Wang et al. 2002, 2010). Our research deviates from much of this literature by focusing on carryover resulting from the *response* to the previous questions as opposed to because of the question content or common structure. We expand the existing literature by exploring how such state dependence may vary over the course of the survey, across response categories, item characteristics, and respondent characteristics.

To get a preliminary sense for the possible degree of state dependence in surveys, Table 1 shows a summary of the responses to a large cross-national marketing research survey consisting of over 100 questions exploring multiple personality and marketing constructs (which we describe in detail in

**Table 1 Repeat Response Frequencies in Study 1**

	Response option (in %)				
	1	2	3	4	5
Probability of response	9.7	22.0	27.9	30.8	9.6
	Previous answer (in %)				
	1	2	3	4	5
Prob. of repeat response	26.8	30.3	38.7	39.8	28.2
Prob. of repeat response: Same construct	49.4	47.8	48.0	57.8	51.1
Prob. of repeat response: Different construct	21.8	26.9	36.7	36.6	24.4

Study 1). To the extent that questions are answered independently, the probability of two subsequent responses being the same should equal the base rate (row 1 in Table 1): 9.7% if the previous answer is a 1, 22% if it is a 2, etc. However, as row 2 shows, the observed percentage of same response to two consecutive questions is always substantially and statistically significantly greater than the percentage predicted by the independence assumption ( $p < 0.05$ ). The difference is most notable for the extreme (1 or 5) response categories. The question is, why?

One possible reason for what may look like “carryover” from one response item to the next is that consecutive questions address the same construct or topic and are therefore answered in a similar manner. To investigate this possibility, we computed the repeat response probability when the two consecutive questions measured the same construct. The results (row 3), unsurprisingly, show an even stronger repeat response rate relative to rows 1 and 2 ( $p < 0.05$ ) when two consecutive questions measure the same construct. However, row 4 provides a more interesting comparison; even when the construct being measured differs between consecutive questions, the repeat response rate is substantially and significantly greater than the base rate for all five response categories ( $p < 0.05$ ).

Although the model-free aggregate findings reported thus far are consistent with and provide suggestive evidence for autocorrelation in responses, other factors may contribute to the observed results. Specifically, the pattern of results in Table 1 could be due to heterogeneity in preferred response (or scale usage heterogeneity), consistent with the well-documented spurious state-dependence effect (Heckman 1981, Keane 1997). For example, what may look like carryover for the extreme responses observed in Table 1 may be due to the presence of yea-sayers and nay-sayers. Therefore, it is important to separate the various possible determinants of response carryover to determine the extent of “true” state dependence in survey response.

The objective of this paper is threefold: (1) to examine the extent of state dependence in survey responses, including how it varies across response categories and items and how it evolves over the course of the survey, (2) to examine whether the degree of state dependence varies as a function of sociodemographics and across countries, and (3) to examine the impact of state dependence on subsequent analyses and inference. To address these issues, we developed a model that explores to what extent observed responses are subject to Markovian state dependence while controlling for item content as well as individual differences in scale usage. The proposed modeling framework integrates dynamic choice models (Paap and Franses 2000, Seetharaman et al. 1999) and dynamic latent variable models (Verhelst and Glas 1993).

Using a large-scale marketing research survey collected across six countries and four continents and a more controlled study involving an experiment to measure the drivers of state dependence and its impact on reliability and validity, we demonstrate the following:

1. State dependence is present in surveys and varies across countries.
2. State dependence is stronger for items that are administered later in a survey, and this effect varies across countries.
3. State dependence varies across response categories. Estimating a common state-dependence parameter across response categories may lead to the erroneous conclusion that there are no state-dependence effects.
4. State dependence is strongest for response categories that are used less frequently.
5. State dependence has deleterious effects on common statistics such as reliability and predictive validity.

These findings provide guidance for survey research on how to analyze and design surveys to minimize the impact of state dependence. The remainder of this article is structured as follows. First, we briefly review the relevant literature on measurement theory and state dependence. Next, we introduce the proposed dynamic latent variable model. Then, we present two empirical applications of the model. The paper ends with conclusions, limitations, and directions for future research.

## Conceptual Foundations

### Measurement Theory

One of the earliest milestones in measurement was common factor theory where items are influenced by a true score (a latent trait) as well as random measurement error. A series of papers in the 1940s by Cronbach (e.g., Cronbach 1946) suggested that systematic, content irrelevant factors also influence

item scores (e.g., acquiescence, a general tendency of individuals to agree with statements independent of content). Hence, to model the response process accurately, an observed score should be decomposed as follows (see also Baumgartner and Steenkamp 2006):

$$\text{Observed score} = \text{True score} + \text{Systematic tendencies} + \text{Random error.} \quad (1)$$

A host of marketing studies have examined the systematic response tendencies: for example, Gruber and Lehmann (1983), Greenleaf (1992), Bradlow and Zaslavsky (1999), Baumgartner and Steenkamp (2001, 2006), Rossi et al. (2001), Wong et al. (2003), de Jong et al. (2008), Rindfleisch et al. (2008), Swain et al. (2008), de Jong et al. (2010), and van Rosmalen et al. (2010). The term “common method bias” is an umbrella term coined to capture various systematic tendencies that might occur during the response process (Podsakoff et al. 2003). We propose including state dependence as an additional systematic tendency, accounting for both varying state dependence throughout the survey and heterogeneity in state dependence across respondents and response categories.

### State Dependence

The notion of state dependence is well established. In economics, Heckman (1981) investigated state dependence to capture the notion that a period of unemployment may causally increase the likelihood of unemployment in the next period. In marketing, state dependence has been used to model brand loyalty (e.g., Guadagni and Little 1983) and to capture advertising persistence as in the Koyck model (Jones and Landwehr 1988). In the context of brand choice, state dependence has been documented across product categories (Seetharaman et al. 1999). State dependence is closely related to the well-documented status quo effect, which postulates that individuals have a strong tendency to remain at the status quo (Kahneman et al. 1991, Samuelson and Zeckhauser 1988).

### Content Carryover vs. State Dependence

State dependence can be viewed as a particular form of context effect. It is important to distinguish state dependence from the related “content” carryover effect (Bickart 1993, Schumann and Presser 1981, Sudman et al. 1996, Tourangeau et al. 2000). Content carryover occurs if respondents give answers that are influenced by beliefs rendered accessible by a previous question. Thus, the answer to a question affects the answer to the next question because of its substantive association, and hence, question order becomes important. Content carryover is particularly relevant when overall evaluations influence

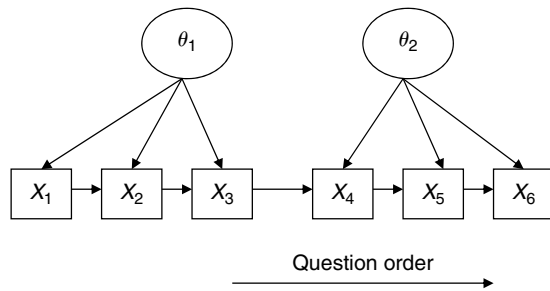
the ratings of specific attributes. Such relationships between holistic evaluations and specific evaluation include part-part and part-whole effects (Schuman and Presser 1981), carryover/backfire effects between general and specific questions about a topic (Bickart 1993), augmented response to unasked questions (Gal and Rucker 2011), and halo effects (Beckwith and Lehmann 1975). In addition, a number of papers consider how grouping and location of items affects reliability (e.g., Bradlow and Fitzsimons 2001, Hertzog and Bachman 1981, Weijters et al. 2009). Indirectly, these papers also shed light on state dependence, as a pattern of similar responses produces high (but potentially spurious) reliability.

Finally, the survey literature has considered “proximity error” (Guilford 1954, Kane 1969), which occurs when adjacent attributes are evaluated more similarly relative to attributes located farther apart. Proximity error has been mainly studied in the employee appraisal research (Zedeck and Blood 1974) and is close in concept to the state-dependence effect examined here.

In this article we study *response carryover* independent of item content or common stimulus and label it “state dependence.” Whereas content carryover refers to the impact of the *content* of the previous questions on the response to the current question, state dependence captures the effect of the *response* to the previous question on the response to the current question. The model-free evidence in Table 1 (row 4) hints that response carryover exists over and beyond content carryover because the probability of repeating a response is significantly higher than chance, even when the two consecutive items do not belong to the same construct. In the empirical applications described later, we utilize item keying (positive versus negative phrasing) and experimental design in which the location and structure of the scale varies across conditions to disentangle these two effects.

A different stream of research that is related to state dependence and our modeling approach is testlet theory (Bradlow et al. 1999; Wainer et al. 2007; Wang et al. 2002, 2010). Testlet theory was developed to capture *within-construct effects* in intelligence tests, wherein items display excess dependence due to common stimulus material. Like our proposed model, testlet theory utilizes a Bayesian Markov chain Monte Carlo approach to disentangle the testlet context effect from scale usage and question content. However, whereas in testlet theory only consecutive items that belong to same stimuli or reading passage should exhibit excess correlation, state dependence implies excess correlation even between consecutive items that are substantially unrelated and do not share a common construct or stimuli. Figure 1 illustrates our model. We expect enhanced correlation

Figure 1 State-Dependence Models



between  $X_1$  and  $X_2$ ,  $X_2$  and  $X_3$ , etc., due to response carryover. Importantly, in our model,  $X_3$  and  $X_4$  will also be correlated because of state dependence, even though they come from two independent constructs ( $\theta_1$  and  $\theta_2$ ).

### Characteristics and Drivers of State Dependence

Several mechanisms could lead to a tendency to repeat the last response independent of content. First, it is physically easier for a respondent to give the same response because he or she does not have to move a pencil/mouse cursor from one item to another. On a more cognitive level, respondents may need to perform costly mental operations to provide an answer. The four-step model of survey response selection (Tourangeau et al. 2000) decomposes the survey response process into comprehension, retrieval, judgment, and response. State dependence allows the respondent to skip most of these steps and reduce the cognitive effort invested in the response by simply selecting the same response given to the previous item (Rindfleisch et al. 2008).

Respondents may also display the opposite effect, i.e., negative state dependence, such that the probability that the response chosen for the next item is identical to the response to the previous item is lower than would be expected based on the next item's content. Negative state dependence may stem from the respondents' need to convey information, appear interesting to the researcher, or follow conversational norms (Grice 1975) that answers to different questions should be different. Alternatively, variety seeking (McAlister and Pessemier 1982, Drolet 2002) could lead to negative state dependence if respondents choose different answers to avoid boredom.

Several factors can affect the valence and magnitude of state dependence. First, state dependence may depend on item characteristics because longer items require more cognitive effort to process than short ones. Second, state dependence may depend on the item sequence. If two consecutive items measure the same construct (e.g., the items in the scale are grouped), this may produce response carryover. Such response carryover for consecutive items may be due to respondents seeing the two items as redundant,

because once information retrieval is initiated, it tends to activate the same beliefs (see Weijters et al. 2009), or due to improved comprehension and reduced cognitive load and error (Tourangeau et al. 2000). In the latter case, carryover is content-related and improves validity. Determining whether validity is improved or not by grouping items is an important question (Bradlow and Fitzsimons 2001). Study 2 will further investigate this important yet vexing issue.

Third, item keying may also influence state dependence. Items can be either positively or negatively keyed. A number of papers in marketing have considered item keying (Swain et al. 2008, Wong et al. 2003). A change in item keying may disrupt routinized responding and encourage the respondent to pay more attention to the item, leading to lower state dependence. This is especially likely when two consecutive items measure the same construct because logically a respondent would move to the other side of the rating scale. Alternatively, changes in item keying may lead to low-quality responses if the respondent does not pay attention to the change. Variation in item keying allows us to disentangle content and response carryover.

Fourth, the degree of state dependence may vary over the course of the survey. Answering questions can be tiring, and the cumulative cognitive effort expended could lead to stronger positive state dependence (see DeSarbo et al. 2004 and Johnson et al. 1990, who demonstrated that response patterns may change during a survey). On the other hand, the desire to provide information and appear interesting and to be a diligent respondent may lead to negative state dependence. These opposing forces may lead to a nonlinear state-dependence effect over the course of the survey.

Fifth, individual and country characteristics may influence the extent of state dependence. Seetharaman et al. (1999) found ample evidence for heterogeneity in state dependence using brand choice data. Observed sources of heterogeneity such as sociodemographics did not explain much of the variation in state dependence in brand choice.

Sixth, in most models of brand choice, the state-dependence effect is assumed to be the same across brands (see Chintagunta 1998 for an exception). In the context of survey response, it is possible that respondents may have different tendencies to be state dependent for different response categories. For example, it is possible that respondents may have stronger tendency to exhibit state dependence to extreme (1 or 5 on a 1-to-5 Likert scale) relative to moderate (2–4) responses, possibly because the mouse cursor can be moved in two ways from the middle and intermediate response categories but only in a single direction for the extreme scale categories.

The purpose of this paper is not to disentangle the exact psychological mechanisms underlying state dependence, but rather to measure its impact. We present a flexible model that captures how state dependence depends on item characteristics, item sequence, individuals, countries, response categories, and the location of the question in the survey. The model is then used to highlight the implications of state dependence for survey-based marketing research.

### Statistical Model

Our proposed model is grounded in item response theory (IRT), which relates the probability of a person responding to an item in a specific manner on the trait that the item is measuring. The proposed model extends existing IRT models by incorporating scale usage heterogeneity as well as state dependence. Like Bradlow and Zaslavsky (1999), our approach captures cognitive dimensions and statistical aspects of survey response. The polytomous response is modeled using a multinomial logit latent variable model (van Rosmalen et al. 2010). This section first presents the basic IRT model and then extends it to incorporate scale usage and state dependence.

Consider a series of  $K$  items, which all measure the same underlying construct. For the sake of exposition, assume that the  $K$  items all measure customer satisfaction. Define  $\theta_i$  as the latent construct score (satisfaction in this case) for respondent  $i$  ( $i = 1, \dots, N$ ). Respondent  $i$ 's utility or value  $U_{i,k,c}$  for response category  $c$  ( $c = 1, \dots, C$ ) for item  $k$  ( $k = 1, \dots, K$ ) can be written as

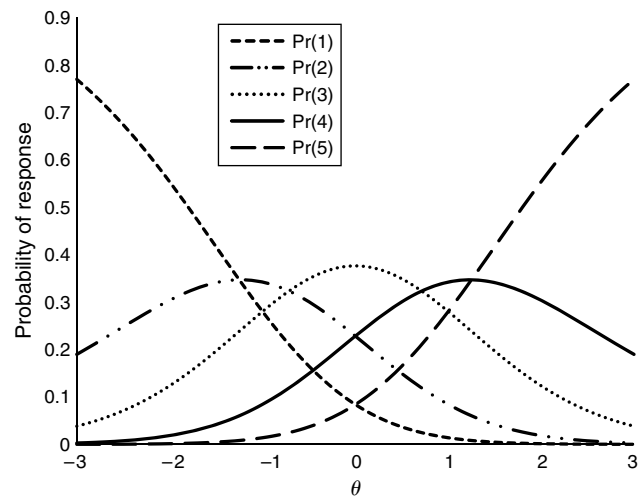
$$U_{i,k,c} = a_{k,c}\theta_i + \varepsilon_{i,k,c}, \tag{2}$$

where  $a_{k,c}$  is a “discrimination” parameter that is specific to both item  $k$  and response category  $c$ , discriminating among people who are high versus low on the latent satisfaction trait  $\theta$ , and  $\varepsilon_{i,k,c}$  is an unobserved random error. Consistent with van Rosmalen et al. (2010), the model does not impose order on the response categories to allow for flexible representation of response styles and question types. Anderson (1984, p. 3) provides another justification for this choice: “There is no merit in fitting an ordered relationship as a routine, simply because the response variable is ordered.” Response category  $c$  is chosen if the utility of response category  $c$  exceeds the utility of all other response categories:

$$X_{i,k} = c \quad \text{if } U_{i,k,c} > U_{i,k,m} \\ \text{for } m = 1, \dots, c - 1, c + 1, \dots, C, \tag{3}$$

where  $X_{i,k}$  is the observed response of person  $i$  on item  $k$ , and the error term  $\varepsilon_{i,k}$  has a Gumbel distribution. Accordingly, the resulting probabilistic item response model is a multinomial logit latent variable model.

Figure 2 Category Response Functions



To illustrate the model, Figure 2 plots examples of category response functions for a five-point scale. In this graph, the discrimination parameters are set as  $a_{k1} = -1.5$ ,  $a_{k2} = -0.7$ ,  $a_{k3} = 0$ ,  $a_{k4} = 0.7$ , and  $a_{k5} = 1.5$ . For negative values of the latent score  $\theta_i$ , the probability of responding 1 is high, whereas other response categories have smaller probabilities. Analogously, for positive values of the latent trait  $\theta_i$ , the probability of responding 5 is high, whereas other response categories have smaller probabilities. The response probabilities of the intermediate categories are larger for values of the latent trait that are close to 0.

### Extended Model

The model in Equation (2) is designed to measure a single construct (e.g., satisfaction). Here, we extend this model to include multiple constructs and disentangle state dependence from scale usage heterogeneity and item content.

**Multiple Constructs.** Surveys often include sets of questions, where each set measures a different construct (Churchill 1979). For instance, one may want to relate overall satisfaction to service quality, in which case at least those two constructs would be measured. Different respondents may have different predispositions to the underlying constructs or traits represented by the questions. We assume that it is known a priori which item corresponds to which construct and that there are multiple items to measure each construct (that is,  $K_d \geq 2 \forall d$ ). The  $K$  items in the survey can thus be broken into  $D$  batteries of questions, where each battery  $d$  of  $K_d$  items represents a construct such that  $K_1 + K_2 + \dots + K_D = K$ . The latent construct scores are contained in the vector  $\theta$ , which is of size  $D$ , and the construct measured by item  $k$  is denoted by  $d(k)$ . Accordingly, we replace  $\theta_i$  in Equation (2) with  $\theta_{i,d(k)}$  in Equation (4), the latent trait

corresponding to item  $k$ . The vector of latent traits is  $\theta_i = (\theta_{i,1}, \dots, \theta_{i,D})' \sim \text{MVN}(\bar{\theta}, \Sigma^\theta)$ , where the intercorrelations among the latent traits are captured in  $\Sigma^\theta$ .

**Scale Usage Heterogeneity.** An additional source of variation in responses comes from heterogeneity in scale usage across respondents. That is, respondents favor particular response categories independent of item content. Although the base model in Equation (2) accounts for heterogeneity in the latent trait ( $\theta_i$ ), it does not account for different baseline tendencies to use certain scale categories. Capturing heterogeneity between respondents is particularly important when one wishes to estimate state-dependence effects because unaccounted-for heterogeneity can lead to spurious state dependence (Heckman 1981, Kahn et al. 1986). To accommodate scale usage heterogeneity, the response category preferences in Equation (2) are allowed to be individual-specific ( $\beta_{i,c}$ ):

$$U_{i,k,c} = \beta_{i,c} + a_{k,c} \theta_{i,d(k)} + \varepsilon_{i,k,c} \quad (4)$$

### State Dependence

State dependence implies that the likelihood of answering response category  $c$  on item  $k$  is influenced by the response to the previous item  $k - 1$  (after controlling for the question content, latent construct, and the respondent's general scale usage). The following extensions to Equation (4) are thus posited:

$$\begin{aligned} U_{i,k,c} &= \beta_{i,c} + a_{k,c} \theta_{i,d(k)} + \gamma_{i,sd,k,c} I(X_{i,k-1} = c) + \varepsilon_{i,k,c} \quad (5) \\ \gamma_{i,sd,k,c} &= \beta_{i,sd,c} + \beta_{i,dyn,1} k + \beta_{i,dyn,2} k^2 + \beta_{i,dyn,3} k^3 \\ &\quad + \beta_{i,keying} CKeying(k) + \beta_{i,condep} I(d(k) = d(k-1)) \\ &\quad + \beta_{i,condep-keying} I(d(k) = d(k-1)) \times CKeying(k) \\ &\quad + \beta_{i,Nchar} NChar(k). \quad (6) \end{aligned}$$

The parameter  $\gamma_{i,sd,k,c}$  in Equation (5) captures carryover from one item to another. The indicator function  $I(X_{i,k-1} = c)$  captures whether the response to the previous item was  $c$ . The indicator function  $I(d(k) = d(k-1))$  equals 1 if the two consecutive items belong to the same construct and 0 otherwise.  $CKeying(k)$  is a dummy variable that equals 1 if item  $k$  and item  $k - 1$  have the same keying (positive or negative) and 0 otherwise.  $NChar(k)$  is the number of characters in item  $k$ . Negative carryover ( $\gamma_{i,sd,k,c} < 0$ ) means that the likelihood for response category  $c$  decreases when category  $c$  was chosen for the previous item, whereas positive carryover ( $\gamma_{i,sd,k,c} > 0$ ) means that the respondent is more likely to choose the same response for subsequent items. We discuss the components of the carryover effect in Equation (6) next.

*Baseline (Response Category-Specific) State Dependence:* Although the majority of the brand choice research pools state dependence across brands, in the context

of survey response pooling, state dependence across response categories may yield a biased picture of the true state-dependence effect. Accordingly, an individual and response category-specific intercept  $\beta_{i,sd,c}$  is included in Equation (6).

*Dynamic State Dependence:* State dependence may vary over the course of the survey as well as across individuals and countries. To allow for dynamic state-dependence effects, state dependence is modeled as a cubic function of the item's location in the survey ( $k$ ), where  $k$  is the ordinal position of the item in the survey. In Equation (6), dynamics are thus captured by three terms:  $\beta_{i,dyn,1} k$ ,  $\beta_{i,dyn,2} k^2$ , and  $\beta_{i,dyn,3} k^3$ . To reduce collinearity, item position is rescaled by the formula  $k' = k/K - 1/2$  (see Liechty et al. 2005 for a similar approach).

*Item Characteristics and Sequence:* To capture the effect of item length, we include the number of characters of the current item  $NChar(k)$  in Equation (6). We use the number of characters instead of the number of words because the number of words is hard to determine for Asian languages.

*Item Keying:* State dependence is likely to be weaker when items change keying over subsequent questions. If item  $k - 1$  is negatively (positively) keyed whereas item  $k$  is positively (negatively) keyed, we predict weaker state-dependence effects (that is, we expect  $\beta_{i,keying} < 0$ ).

*Construct Dependence:* In addition to the general tendency to give similar (or different) answers to consecutive items (state dependence), there could be an additional "bump" if two consecutive items belong to the same scale (Bickart 1993, Bradlow and Fitzsimons 2001, Weijters et al. 2009). We call this construct dependence. Note that the response consistency as a result of the construct, dependence component in Equation (6) could be caused by both content and response carryover. To incorporate this effect, Equation (6) is augmented with an additional term,  $\beta_{i,condep} I(d(k) = d(k-1))$ .

Furthermore, the effect of item keying is likely to interact with construct dependence. If keying changes and the two consecutive items measure the same construct, respondents should theoretically switch to the opposite side of the rating scale. To capture this effect, we include in Equation (6) an interaction term  $I(d(k) = d(k-1)) \times CKeying(k)$ . Note that the interaction between keying and construct dependence can disentangle response state dependence from content carryover effects. Content carryover would imply that change in keying over consecutive items of the same construct will produce different responses whereas response state dependence would produce the same response to consecutive items with alternating keying.

*Individual Characteristics:* The state-dependence parameters and the respondent's response category-specific intercepts are allowed to depend on a set of individual-specific variables via a hierarchical Bayesian specification. These covariates are captured in the vector  $Z_i$ :

$$\beta_i = (\beta_{i,1}, \dots, \beta_{i,C}, \beta_{i,sd,1}, \dots, \beta_{i,sd,C}, \beta_{i,dyn,1}, \beta_{i,dyn,2}, \beta_{i,dyn,3}, \beta_{i,keying}, \beta_{i,condep}, \beta_{i,condep-keying}, \beta_{i,NChar})' \\ \sim \text{MVN}(\tilde{\beta}' Z_i, \Sigma^\beta). \quad (7)$$

Overall, the extended model in Equations (5)–(7) simultaneously captures: (1) individual response tendencies/scale usage heterogeneity, (2) question content (construct), (3) baseline response category-specific state dependence, (4) state-dependence variation over the course of the survey, (5) state-dependence variation across individuals, (6) the effect of item length on state dependence, (7) the effect of item keying on state dependence, (8) construct dependence, and (9) the interaction of item keying and construct dependence.

### Identification

Several identification restrictions are necessary. First, the scale of the latent (trait) variables needs to be set. We set the last element of the vector  $\theta$  to 0. Second, the variance parameters of all latent constructs are set equal to 1 to fix the scales. Third, the discrimination parameters for each item are subject to the following restriction:  $\sum_c a_{k,c} = 0 \forall k$ , and  $a_{k,c} > 0$  ( $< 0$ ) for positively (negatively) keyed items. Fourth, the response style parameters have to satisfy  $\sum_c \beta_{i,c} = 0$ .

It is necessary to make sure the survey design and model allow one to properly separate the various behavioral effects (item content, scale usage, state dependence, and other context effects). The following conditions are helpful for disentangling the various effects.

*Condition 1.* At least a subset of items in the survey needs to be heterogeneous and diffuse in terms of content (that is, less than perfect interitem correlation among items that measure different constructs for at least a few constructs). If all across-construct items are highly correlated, it is difficult to disentangle item content from systematic response tendencies (Baumgartner and Steenkamp 2001). This condition was satisfied in Studies 1 and 2 and is often true for commercial marketing research studies.

*Condition 2.* Multi-item measures are useful in separating content from response tendencies. Our surveys consist of multi-item measures.

*Condition 3.* A statistical model that explicitly captures the various effects is needed.

A question that may arise is what in the data allows us to uniquely separate the effects of respondent's

construct score ( $\theta_{i,d(k)}$ ), construct dependence carryover ( $\beta_{i,condep}$ ), and scale usage heterogeneity ( $\beta_{i,c}$ ). Intuitively, the fact that the state-dependence parameters are not construct-specific and are estimated across many items and constructs (some of them with less than perfect interitem correlation by design) allow us to disentangle these effects. Three main variations in the data allow us to identify state dependence: (1) within-construct "runs" of the same response over and beyond what is observed in the population or expected by chance (e.g., a response sequence of 2, 2, 2, 3, 3, 3 versus 2, 3, 3, 2, 3, 2), (2) carryover of the same response from the last item of one construct to the first item of the next construct over and beyond what is observed in the population or expected by chance, and (3) identical responses to consecutive items with alternating keying that belong to the same construct. Variation on any of these three dimensions in any construct lead to identification of state dependence for all constructs. We provide an intuitive example that demonstrates how different response patterns enable identification of state dependence in Web Appendix A (at <http://dx.doi.org/10.1287/mksc.1120.0722>).

Because we assume that state dependence is not construct-specific, less than perfect within- and across-construct correlations for at least some constructs suffice for identification. At the extreme, a survey in which the respondents give the same response to all items would be totally useless. This identification restriction is similar to the requirements for detection of common method bias. Researchers generally recommend the use of marker items that are theoretically unrelated to the focal measures in the study to detect such biases (Lindell and Whitney 2001).

To examine the ability of our model to uniquely separate these effects, we conducted a simulation study in which we generated survey data similar to the data from our empirical applications. In the simulation we also increased the degree of within- and between-construct item correlations to examine boundary condition for identification. The simulation confirms that we are able to uniquely recover all the model's parameters and in particular those for state dependence, construct traits, and scale usage heterogeneity. We report details of the simulation in Web Appendix A. Additionally, Study 2 further facilitates identification through an experimental design.

### Estimation

The model is estimated using Markov chain Monte Carlo methods. Estimation is computationally very intensive because of the high-dimensional nature of the model. The full set of prior and posterior distributions is available in Web Appendix B. A Metropolis-Hastings algorithm is used to sample  $\theta_i$  and  $\beta_i$ .

The full conditionals of the discrimination parameters are log-concave and are sampled using adaptive rejection sampling (ARS; see Gilks and Wild 1992). ARS techniques sample parameter values from an envelope function of the conditional distribution. The correlation matrix  $\Sigma^\theta$  is sampled using a Metropolis Hit-and-Run algorithm (e.g., Ansari and Jedidi 2000). The first 20,000 iterations were used as “burn-in,” and the last 2,500 thinned 1 in 2 were used to calculate the posterior distributions. Convergence was assessed by visually examining the trace plots and by using Gelman and Rubin’s (1992) convergence diagnosis.

## Empirical Applications

In this section we apply the model to survey-based data sets. Study 1 employs a cross-national survey and examines the existence of state dependence in survey response and how it varies across countries. Study 2 uses an experimental design to check the robustness of the state-dependence phenomenon and to test the impact of state dependence on reliability and predictive validity.

### Study 1: Cross-National Survey

**Data.** The data come from a multinational marketing research study. Two global marketing research agencies, the GfK Group and Taylor Nelson Sofres (TNS), collected self-reported survey data in six major countries: the United States, Germany, Russia, China, Japan, and Brazil. The survey focused on several aspects of personality, well-being, and consumer behavior. The sample in each country was representative of the total population in terms of region, age, education, and gender. For countries with high Internet penetration (the United States, Germany, Japan), a completely Web-based survey was used. In countries with low Internet penetration (China, Brazil, Russia), data were collected by computer-assisted mall intercept surveys in multiple regions/locations. The mall intercept surveys were also Web-based and filled in by the respondent on a laptop at the point of “intercept” (e.g., Internet cafés, subsidiaries of offices, and test halls for product tests). The number of respondents per country varied between 396 (China) and 640 (Germany). The total number of respondents was 2,903.

The questionnaire included 100 five-point Likert scale (1 = strongly disagree, 5 = strongly agree) questions, measuring the Big Five personality dimensions, life satisfaction, positive and negative affect, and several marketing scales. Of the 100 questions, 21 were negatively keyed. For each personality factor, three highly loading positively and negatively worded items were selected based on the cross-cultural study

of Benet-Martinez and John (1998) (openness to experience included a fourth item). Affective well-being was assessed with items from the Affectometer 2 scale (Kamman and Flett 1983). Positive and negative affect were measured with 10 adjectives each. Finally, life satisfaction was measured with the five-item Satisfaction with Life Scale developed by Diener et al. (1985).

The marketing-related constructs measured were susceptibility to normative influence (Bearden et al. 1989), innovativeness (Steenkamp and Gielens 2003), deal proneness (Lichtenstein et al. 1995), nostalgia (Holbrook and Schindler 1994), quality consciousness (Steenkamp 1989), material success (Richins and Dawson 1992), environmental consciousness (Grunert and Juhl 1995), consumer ethnocentrism (Shimp and Sharma 1987), and health consciousness (Baumgartner and Steenkamp 2001). All items were taken from previously validated scales. A subset of the items in an established scale was used in most cases because of time constraints and the desire to avoid overburdening respondents (Burisch 1984). The average absolute within-construct interitem correlation for this data set is 0.377 with a minimum of 0.16, a maximum of 0.69, and a standard deviation of 0.135. The average absolute between-construct interitem correlation is 0.135, with a minimum of 0.001, a maximum of 0.548, and standard deviation of 0.11. These correlations indicate that we have a heterogeneous and diffuse set of items.

In addition, the questionnaire included sociodemographic questions such as age (measured in years), gender (1 for women, 0 for men), and education (ranging from “no formal education” to “university”). We used these questions as covariates in Equation (7). The questionnaire was developed in English and then translated into the appropriate local languages by professional agencies, using back-translation to ensure accuracy.

For some of the scales, the items in the scale were dispersed across the survey, whereas for other scales, the items were grouped or partially grouped, thus allowing us to investigate construct dependence. For this data set, it was not possible to estimate the interaction effect between construct dependence and change in item keying ( $\beta_{i, \text{condep-keying}}$ ) on state dependence because there are no instances where two consecutive items that belong to the same scale change keying. Hence, for the current data set, we dropped the interaction term from Equation (6) while estimating the main effect of change in keying and the construct-dependence effect. Study 2 allows for investigation of this interaction effect.

**Results.** We first compare the fit of our proposed model to a benchmark model that does not account for state dependence but accounts for scale usage heterogeneity, latent traits, and observed heterogeneity



(the model described in Equations (5) and (6) setting  $\gamma_{i, sd, k, c}$  to 0). The log marginal density (LMD) for the model with state dependence is equal to  $-272,154$  relative to a log marginal density of  $-308,122$  for the model without state dependence, suggesting that substantial state dependence exists in the responses.

To further examine the existence of state dependence in the data, we compared the average repeat response rates to subsequent questions in Table 1 with repeat response rates after randomizing the order of questions prior to calculating the repeat response rate. Increased repeat response rate in the randomized data set could only be attributed to common latent traits across questions or scale usage heterogeneity. The average repeat response rate for the original data set was 32.8%, relative to 24.5% for the randomized data set. Thus, although the randomized repeat response rates are higher than chance (20%), there is a substantial and significant increase in repeat response rates when one accounts for the order of questions in the questionnaire (state dependence). Consecutive items that did not belong the same construct have a 29.3% repeat response rate for the original data set relative to a 24.4% repeat response rate for the randomized data set, suggesting that state dependence goes beyond common construct and content carry-over effects.

Having established that state dependence is present in the data, we describe response category preferences across countries and respondents' demographics. In addition, we examine the magnitude and statistical significance of the state-dependence parameters across countries.

*Response Category Preferences:* Response style did not vary substantially by age, gender, and education. The left entry in each cell of Table 2 is the number of times respondents in each country use each response category (often called zero-order probabilities in the brand choice literature). A limitation of the model-free summary in Table 2 is that it does not account for alternative drivers of response (e.g., state dependence). Accordingly, we compare the results in Table 2 with the posterior mean estimates for the scale

**Table 2** Response Frequencies and Repeat Probabilities Across Countries in Study 1

	Response option (Probability of response/repeat probability (in %))				
	1	2	3	4	5
Brazil	7.6/27.6	22.3/30.7	21.5/33.4	36.7/47.5	11.9/30.8
China	3.9/27.3	19.3/30.7	32.2/45.6	35.8/45.2	8.9/30.5
Germany	11.9/27.0	22.6/28.3	26.6/31.9	31.4/37.2	7.6/20.2
Japan	8.7/26.4	25.1/33.1	36.0/44.8	24.7/31.9	5.5/19.2
Russia	14.5/29.1	19.1/29.6	20.9/35.1	27.0/37.1	18.5/35.1
United States	10.4/23.9	22.8/29.5	28.9/36.9	29.8/35.8	8.0/23.1

**Table 3** Posterior Mean Estimates of Scale Usage in Study 1

	$\bar{\beta}_1$	$\bar{\beta}_2$	$\bar{\beta}_3$	$\bar{\beta}_4$	$\bar{\beta}_5$
Intercept	<b>-2.718</b>	<b>0.232</b>	<b>1.119</b>	<b>1.183</b>	<b>-1.453</b>
Age	0.000	0.001	0.000	<b>0.003</b>	<b>-0.004</b>
Gender	<b>0.075</b>	-0.014	-0.028	<b>-0.039</b>	0.003
Education	-0.021	0.012	0.014	0.014	-0.020
Germany	0.174	<b>-0.258</b>	<b>-0.369</b>	-0.247	-0.141
Russia	<b>0.864</b>	<b>-0.472</b>	<b>-0.856</b>	<b>-0.502</b>	<b>0.978</b>
Brazil	-0.313	<b>-0.391</b>	<b>-0.795</b>	<b>-0.247</b>	<b>0.293</b>
China	<b>-1.098</b>	<b>-0.204</b>	-0.021	<b>0.203</b>	0.220
Japan	0.145	0.026	0.015	<b>-0.431</b>	<b>-0.517</b>

Notes. Estimates in bold indicate "significant at 5%" in frequentist terms. The United States is omitted. Thus, all country effects should be interpreted relative to the United States.

usage parameters ( $\beta_{i,1}, \dots, \beta_{i,5}$ ) across countries (see Table 3).

Across the six countries, respondents tended to avoid the extremes (categories 1 and 5) of the rating scale (de Jong et al. 2008). Relative to the respondents in the United States (the baseline country in Table 3 and last row in Table 2), Russian respondents tend to select the extreme response categories 1 and 5 much more frequently ( $\bar{\beta}_{1, Russia} = 0.864, \bar{\beta}_{5, Russia} = 0.978$ ) and categories 2, 3, and 4 much less often ( $\bar{\beta}_{2, Russia}, \bar{\beta}_{3, Russia},$  and  $\bar{\beta}_{4, Russia}$  are all negative and significant).

German respondents responded mostly similarly to U.S. respondents, except for a slightly lower preference for categories 2 and 3. Brazilian respondents, on average, had a slightly higher tendency to use the positive part of the scale.

Japanese and Chinese response patterns were remarkably dissimilar. Chinese respondents were the most averse (among the six countries studied) to choosing disagreement categories 1 and 2 ( $\bar{\beta}_{1, China} = -1.098, \bar{\beta}_{2, China} = -0.204$ ). The Japanese respondents, on the other hand, were much less likely than the U.S. and Chinese respondents to choose response categories 4 and 5 ( $\bar{\beta}_{4, Japan} = -0.431, \bar{\beta}_{5, Japan} = -0.517$ ); i.e., they were less likely to agree with statements. Therefore, the purported Asian "acquiescence bias" does not seem to extend to Japan. The significant and substantial difference between the Japanese and Chinese response tendencies suggests that cross-cultural studies that group together Asian respondents across countries or generalize the results from one Asian country to the whole region may be misleading.

**State-Dependence Parameters.** The right entry of each cell in Table 2 presents the model-free repeat rates across countries. The state-dependence posterior means from Equation (6) are presented in Table 4. The first-order switching probabilities for each country are presented in Web Appendix C.

*Baseline State Dependence:* First, following common practices in brand choice models, we estimated a model with constant state dependence across

**Table 4** Posterior Mean Estimates of State Dependence in Study 1

	$\bar{\beta}_{sd,1}$	$\bar{\beta}_{sd,2}$	$\bar{\beta}_{sd,3}$	$\bar{\beta}_{sd,4}$	$\bar{\beta}_{sd,5}$	$\bar{\beta}_{dyn,1}$	$\bar{\beta}_{dyn,2}$	$\bar{\beta}_{dyn,3}$	$\bar{\beta}_{keying}$	$\bar{\beta}_{condep}$	$\bar{\beta}_{NChar}$
Intercept	<b>0.293</b>	<b>-0.135</b>	<b>-0.157</b>	-0.091	<b>0.429</b>	-0.030	<b>0.589</b>	<b>0.882</b>	<b>-0.110</b>	<b>0.704</b>	<b>0.365</b>
Age	-0.0005	0.001	0.001	0.002	<b>0.004</b>	0.0001	-0.007	0.000	<b>-0.0009</b>	<b>0.002</b>	-0.0004
Gender	0.015	0.011	0.014	0.010	-0.110	0.006	-0.228	0.026	0.010	0.013	0.033
Education	-0.020	<b>-0.041</b>	-0.020	-0.001	-0.002	0.008	-0.001	0.002	<b>0.002</b>	<b>0.043</b>	<b>0.040</b>
Germany	0.106	0.027	-0.096	0.037	-0.040	<b>-0.210</b>	-0.179	<b>1.041</b>	<b>0.143</b>	<b>-0.172</b>	<b>-0.234</b>
Russia	0.045	<b>0.199</b>	<b>0.388</b>	<b>0.149</b>	0.053	0.154	<b>-0.640</b>	0.575	<b>0.234</b>	<b>-0.157</b>	<b>-0.204</b>
Brazil	<b>0.702</b>	<b>0.220</b>	<b>0.255</b>	<b>0.314</b>	<b>0.368</b>	<b>0.196</b>	-0.221	0.124	0.054	<b>-0.358</b>	<b>-0.284</b>
China	<b>1.071</b>	<b>0.392</b>	<b>0.322</b>	0.071	<b>0.375</b>	<b>0.292</b>	0.202	-0.173	<b>0.159</b>	<b>-0.291</b>	<b>-0.251</b>
Japan	<b>0.302</b>	<b>0.235</b>	<b>0.199</b>	0.153	0.209	<b>0.107</b>	<b>-0.681</b>	<b>-0.983</b>	-0.036	-0.021	<b>-0.423</b>

Notes. Estimates in bold indicate “significant at 5%” in frequentist terms. The United States is omitted. Thus, all country effects should be interpreted relative to the United States.

response categories. This model resulted in insignificant state-dependence effects for the United States and Germany,<sup>1</sup> suggesting that state dependence may not exist in these countries. However, estimating a model with response option-specific effects uncovered state-dependence parameters that vary substantially and significantly across response categories (see columns 2–6 in Table 4). State dependence at the response option level may have implications for the order in which item blocks are presented in a survey—if a particular item block leads people to mark extreme responses, this may have an impact on their responses in later item blocks.<sup>2</sup>

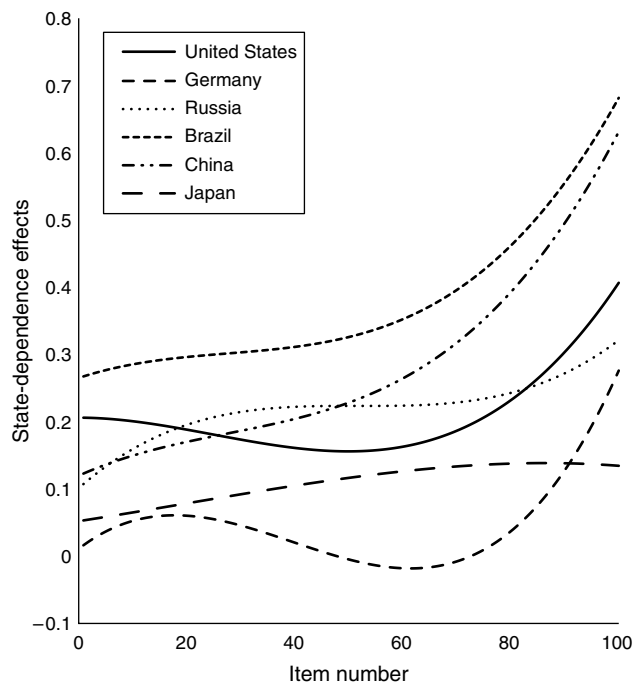
In the United States there is positive baseline state dependence for the extreme response categories 1 and 5 ( $\bar{\beta}_{1, sd} = 0.293, \bar{\beta}_{5, sd} = 0.429$ ) and negative (but smaller in magnitude) state dependence for categories 2 and 3 ( $\bar{\beta}_{2, sd} = -0.135, \bar{\beta}_{3, sd} = -0.157$ ). The result of stronger baseline state-dependence effects for the extreme categories relative to intermediate ones is robust across countries, although the magnitudes vary. For the other countries, except Germany, the baseline state dependence for all response categories is positive. In China, state dependence is stronger than other countries, in particular for the “strongly disagree” option ( $\bar{\beta}_{1, sd, China} = 1.071$ ). Interestingly, respondents across the six countries tend to be more state dependent to the response options for which they have a lower tendency to respond. For example, although the Chinese have the lowest tendency to choose the “strongly disagree” option, once they choose it, they are the most likely to continue choosing it. Similarly, although the Russian respondents are least likely to choose response option 3, once

these respondents choose this option, they have the highest level of state dependence for that category. Baseline state dependence did not vary substantially by age, gender, and education.

*State-Dependence Dynamics:* In all countries except Japan, there was a nonlinear pattern of state dependence over the course of the survey (see columns 7 to 9 in Table 4). Figure 3 plots the state-dependence effect (averaged across response categories by their relative weight) as a function of item position.

In most countries state dependence rises as the survey progresses. This result is consistent with the notion that respondents’ cognitive and mental resources deplete as the cumulative mental effort expended increases (Levav et al. 2010). The implications for researchers are clear. Surveys should not be too long, as state dependence may rise quickly later in the survey, and important statistics could

**Figure 3** State-Dependence Dynamics in Study 1



<sup>1</sup> The pooled estimated state-dependence baseline (United States) parameter was  $-0.045$ , with a 95% credible interval (CI) =  $[-0.088, 0.005]$ . The state-dependence parameters for the other countries were  $-0.028$  with 95% CI =  $[-0.088, 0.039]$  (Germany),  $0.213$  with 95% CI =  $[0.143, 0.294]$  (Russia),  $0.295$  with 95% CI =  $[0.216, 0.379]$  (Brazil),  $0.226$  with 95% CI =  $[0.139, 0.317]$  (China), and  $0.179$  with 95% CI =  $[0.096, 0.256]$  (Japan).

<sup>2</sup> We thank an anonymous reviewer for this suggestion.

be affected. Japan is a clear exception; apparently Japanese respondents are not affected by length of the survey as much as respondents in other countries.

*Item Characteristics:* In general, the more characters the current item has, the more likely the response to that item will be the same as the previous response.<sup>3</sup> This result is consistent with the hypothesized positive relationship between cognitive effort and state dependence. The effect is most pronounced in the United States ( $\bar{\beta}_{NChar} = 0.365$ , last column in Table 4). In terms of sociodemographics, more educated respondents are affected more by item's length than respondents with lower education. Although the education effect is small in magnitude, it suggests that educated respondents are more easily bored and/or less motivated to fill out the survey, and thus they resort to simplifying heuristics when answering questions. More research is necessary to investigate this in more detail. The parameter corresponding to  $CKeying(k)$  is negative but relatively small in magnitude; we investigate this issue further in Study 2.

*Construct Dependence:* Consistent with the model-free evidence in Table 1, the construct-dependence effect is highly significant in all countries and strongest in the United States ( $\bar{\beta}_{condep} = 0.704$ ).

*Mode of Administration:* One possible confound with our cross-national analysis is that mode of administration varied across countries. Although the survey was administered using a Web-survey in all countries, some countries (Japan, Germany, and the United States) used a Web-based Internet panel survey whereas others (Brazil, China, and Russia) used a Web-based mall intercept survey. To test for a mode-of-administration effect, we reran our model with mode of administration (Internet panel versus computer-based mall intercept) instead of country effects (country fixed effects and mode of administration are not jointly identified). The analysis suggests that mall intercept respondents were slightly more likely to exhibit state dependence for the extremes and were less influenced by construct dependence, length of the survey, and the number of characters, although qualitatively, the pattern of the state-dependence effects remains. We note that because of the small number of countries in each mode of administration, the mode of administration is confounded by country-specific effects. The fit of the model with mode-of-administration fixed effects (LMD = -274, 452) is worse than the fit of the model with country-specific fixed effects (LMD = -272, 154), which suggests that country differences, over and beyond the mode of administration, exist in our data.

<sup>3</sup> To investigate whether state dependence becomes more pronounced after a certain threshold point, we also estimated a nonlinear function of the number of characters. However, the nonlinear model was rejected in favor of the linear model.

In sum, Study 1 demonstrates that state-dependence effects at the response-level exist in the investigated survey. Furthermore, the degree of state dependence varies across countries. Not accounting for variation in the state-dependence effect across response categories can lead to the misleading inference of a lack of state dependence. State-dependence effects tend to be strongest for the response categories chosen least frequently and for items placed later in the survey.

## Study 2

Although the statistical model allows us to estimate the influence of state dependence on statistics that are of interest to marketing managers, such as reliability and validity, a cleaner approach is to directly manipulate factors that relate to state dependence and assess their effects on these important statistics. Additionally, state-dependence dynamics found in Study 1 suggest that longer surveys may show stronger state-dependence effects. Finally, Study 1 does not allow for investigation of the “construct-dependence  $\times$  change in keying” effect due to limited co-occurrence of these two factors in the study. Thus, the goal of the second study was fivefold: (1) to use a “clean” experimental design to assess the factors that affect state dependence, (2) to investigate state-dependence dynamics for a longer survey, (3) to estimate the “construct-dependence  $\times$  change in keying” effect, (4) to investigate how state dependence affects reliability estimates, and (5) to investigate how state dependence affects predictive validity.

**Study Design.** Study 2 used a survey consisting of 157 five-point Likert scale items (1 = strongly disagree, 5 = strongly agree). As in Study 1, the scales include the Big Five personality dimensions, life satisfaction, positive and negative affect, and a host of marketing scales drawn from the scales used in Study 1 and from Bearden and Netemeyer (1999). The number of items per scale ( $K_i$ ) ranged from 4 to 10. The results from Study 1 suggest a number of factors that may influence state dependence: position in the survey, item length, item-keying sequence, and grouping versus dispersing items that measure the same construct. Because the effects for item length were modest, in Study 2 we manipulate item position, keying sequence, and item grouping. We also investigate the interaction between the keying sequence and construct dependence to shed more light on the underlying response process.

The focal scale was material success (taken from Richins and Dawson 1992), measured by four items. The experimental design randomly assigned respondents to one of eight conditions. In these conditions, the *location* (toward the beginning: questions

20–40; toward the end: questions 115–135), *grouping* (grouped: all the scale items appeared one after the other; ungrouped: there were two filler items from a different scale between each of the focal scale’s items), and *order* of items within the focal scale. Changing the order of items within a scale allows investigating variation in *keying sequence* of the items. The item-keying sequence for the balanced scale with four items was either “alternating” (PK-NK-PK-NK) or a “single-change” (e.g., the sequence changes keying only once, PK-PK-NK-NK), where PK denotes a positively keyed item and NK denotes a negatively keyed item.<sup>4</sup> The experimental design allows us to test how state dependence and the reliability of the material success scale varies based on the location of the scale in the survey, scale grouping, and the item-keying sequence.

*State Dependence and Predictive Validity:* To investigate the effects of state dependence on predictive validity, Study 2 employed a two-wave study. One of the scales used in the first wave is the consumer innovativeness scale (Steenkamp et al. 1999). Consumer innovativeness is defined as “the predisposition to buy new and different products and brands rather than remain with previous choices and consumption patterns” (Steenkamp et al. 1999, p. 56). Six items were used to measure innovativeness, of which three were positively and three were negatively keyed. The *location* (toward the beginning: starting at question 11; toward the end: starting at question 130) of the innovativeness scale and the *grouping* of the items were manipulated.

Predictive validity was assessed as the correlation between consumer innovativeness measured in the first wave of the survey and the respondent’s reported purchase of new products (Baumgartner and Steenkamp 1996) measured in the second wave, one week after the first wave. The adoption data were collected a week after the first wave to temporally separate the predictive measure from the scale scores. In the second wave, the same respondents were asked to report whether they had adopted products that were considered innovative at the time of the study (an Apple iPad, an e-book reader such as Kindle, high-definition television service, a Blu-ray DVD player, 3D TV, teeth-whitening strips, satellite radio, an Apple iPhone 4G or a similar smartphone, an Apple MacBook computer, Internet telephone (Voice over Internet Protocol), a Nintendo Wii game or Xbox Kinect console, Windows 7, a digital photo frame, and an iRobot that vacuums and cleans). We expect that respondents high in innovativeness will own more of

these new products than those low in innovativeness (see Baumgartner and Steenkamp 1996).

**Data Collection and Model Results.** A random sample of 455 respondents representative of the U.S. population was recruited by Qualtrics, a professional marketing research firm. Respondents were paid \$7 for their participation in wave 1 and \$5 for their participation wave 2 of the study. Of the 455 respondents who completed wave 1, 347 respondents also completed wave 2 (one week later). Attrition did not significantly vary across experimental cells. The average within-construct absolute interitem correlation for the data set in wave 1 is 0.397, and the average between-construct absolute interitem correlation is 0.124, indicating that it is a heterogeneous and diffuse set of items. Table 5 presents the response frequencies and repeat rates for subsequent items in this survey. The scale usage patterns and repeat rates strongly resemble those from Study 1 (see Table 1). That is, response option 4 (“agree”) is used most often, followed by response option 3 (“neither agree nor disagree”). As in Study 1, response categories 1 and 5 were chosen least frequently. Similarly, the repeat rates replicate those from Study 1.

Because of the nature of the design, we can investigate how item keying affects repeat rates (see the last two rows of Table 5). Not surprisingly, repeat response is highest when two subsequent items belong to the same construct and have the same keying. More interestingly, when keying changes between two subsequent items of the same construct, we should not expect repeat responses (apart for response option 3). For example, respondents who answered “4” to an item should theoretically answer “2” to the next, reverse-coded item of the same construct. However, in 32.5% of the cases, they gave the same response. Similarly, for response Option 5, the repeat response of reverse-coded, same-construct subsequent

**Table 5 Repeat Response Frequencies in Study 2**

	Response option (in %)				
	1	2	3	4	5
Probability of response	10.9	17.8	28.0	31.8	11.5
	Previous answer (in %)				
	1	2	3	4	5
Prob. of repeat response	35.4	29.4	46.0	44.4	41.1
Prob. of repeat response: Same construct	52.6	39.9	50.2	53.1	51.9
Prob. of repeat response: Different construct	26.0	23.2	43.4	38.8	33.4
Prob. of repeat response: Same construct, same keying	66.1	49.5	51.4	60.7	62.4
Prob. of repeat response: Same construct, different keying	13.0	13.3	46.9	32.5	21.6

<sup>4</sup> Note that the keying sequence is only relevant if items of a scale are grouped. If items of a scale are dispersed, the filler items determine the keying effects.

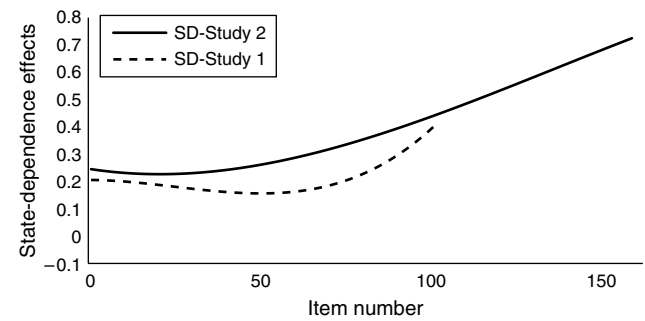
**Table 6** Scale Usage and State-Dependence Parameters in Study 2

	Model M <sub>1</sub>		Model M <sub>2</sub>	
	Posterior mean	95% CI	Posterior mean	95% CI
$\bar{\beta}_1$	-2.241	[-2.500, -1.975]	-2.303	[-2.507, -2.101]
$\bar{\beta}_2$	-0.528	[-0.748, -0.320]	-0.261	[-0.371, -0.151]
$\bar{\beta}_3$	0.144	[-0.076, 0.344]	0.452	[0.341, 0.572]
$\bar{\beta}_4$	0.280	[0.074, 0.476]	0.607	[0.500, 0.715]
$\bar{\beta}_5$	-1.688	[-1.963, -1.415]	-1.390	[-1.591, -1.21]
$\bar{\beta}_{sd,1}$	0.969	[0.865, 1.077]	0.846	[0.717, 0.968]
$\bar{\beta}_{sd,2}$	0.229	[0.156, 0.297]	0.172	[0.098, 0.247]
$\bar{\beta}_{sd,3}$	0.300	[0.217, 0.387]	0.224	[0.133, 0.315]
$\bar{\beta}_{sd,4}$	0.212	[0.144, 0.284]	0.185	[0.112, 0.264]
$\bar{\beta}_{sd,5}$	0.867	[0.743, 0.983]	0.859	[0.728, 0.983]
$\bar{\beta}_{dyn,1}$	0.460	[0.326, 0.585]	0.434	[0.311, 0.5645]
$\bar{\beta}_{dyn,2}$	0.611	[0.545, 0.687]	0.625	[0.561, 0.697]
$\bar{\beta}_{dyn,3}$	-0.532	[-0.769, -0.273]	-0.511	[-0.700, -0.184]
$\bar{\beta}_{keying}$	-0.476	[-0.528, -0.422]	-0.071	[-0.139, -0.005]
$\bar{\beta}_{condep}$	0.440	[0.391, 0.485]	0.460	[0.412, 0.519]
$\bar{\beta}_{condep-keying}$			-0.496	[-0.581, -0.403]
$\bar{\beta}_{NChar}$	0.143	[0.014, 0.269]	0.139	[0.008, 0.265]

items is significantly higher than what would be expected by chance based on response frequencies. This pattern is consistent with response carryover but not with content carryover. This finding also explains why the factor structure of mixed scales is often distorted by the presence of reverse-coded items (Swain et al. 2008).

Next, the statistical model was calibrated on the new data set. The scale usage and state-dependence parameters are provided in Table 6. Two models are calibrated: one similar to the model used in Study 1 without the interaction effect between construct dependence and change in keying (Model M<sub>1</sub>) and the other (Model M<sub>2</sub>) with this interaction effect. The estimation results for Model M<sub>1</sub> in Table 6 are consistent with the results in Study 1 (see Tables 3 and 4). The extreme response categories 1 (“strongly disagree”) and 5 (“strongly agree”) are the least likely to be chosen (lowest intercepts,  $\bar{\beta}_1 = -2.241$  and  $\bar{\beta}_5 = -1.688$ ) and display the strongest state dependence ( $\bar{\beta}_{1,sd} = 0.969$  and  $\bar{\beta}_{5,sd} = 0.867$ ). Figure 4 plots how state dependence changes as the survey progresses and compares the state-dependence dynamics to the dynamics found in Study 1. As in Study 1, state dependence is clearly stronger at the end of the survey. Furthermore, the pattern of state-dependence dynamics is quite consistent across the two surveys.<sup>5</sup>

<sup>5</sup> Note that the data in Study 2 were collected by a different company (Qualtrics) and a number of years later than the data used in Study 1 (collected by GfK). There could be a data provider effect that explains some of the differences observed across studies. For instance, we do not know exactly how Qualtrics and GfK

**Figure 4** State-Dependence Dynamics in Studies 1 and 2

Item length, item grouping, and change in keying are all significant and are directionally consistent with the effects found in Study 1. Overall, the results of Study 2 are consistent with those of Study 1, providing convergent validity evidence for our results.

In Model M<sub>2</sub>, the construct-dependence  $\times$  change in keying effect is negative and significant ( $\bar{\beta}_{condep-keying} = -0.496$ ). Thus, in line with our expectations, if two consecutive items belong to the same construct, reverse coding leads to negative carryover. Note that other terms in the equation may still produce positive overall carryover, even for consecutive reverse-coded items that belong to the same construct.

**State Dependence and Reliability.** To examine the impact of *location*, *grouping*, and *order* of the items on reliability, we computed Cronbach’s alpha (Nunnally and Bernstein 1994) and average interitem correlations by condition for the balanced material success scale (see Table 7). To test for significance of the differences in Cronbach’s alpha/average interitem correlation between the experimental conditions, we use a bootstrap approach to compute tail-area probabilities. That is, we repeatedly sample 3,000 times from the different experimental cells, and for each bootstrap sample, we compare the Cronbach’s alpha values. Averaging the  $I(\alpha_A > \alpha_B)$  over bootstrap iterations yields the tail-area probability, where  $\alpha_A$  and  $\alpha_B$  are the Cronbach’s alphas for a pair of experimental cells A and B, and  $I(\cdot)$  is an indicator function. The tail-area probabilities indicate that grouping has a significant main effect on reliability. When items are grouped, reliability for the balanced scale is lower than when items are dispersed ( $\Pr(\alpha_{grouped} > \alpha_{dispersed}) = 0.002$ ). The main effect for location is also significant ( $\Pr(\alpha_{end} > \alpha_{begin}) = 0.04$ ), and there is a marginally significant interaction between grouping and location ( $\Pr(\alpha_{end} > \alpha_{begin}) = 0.051$  for grouped items).

Keying sequence has little effect on response at the start of the survey because if attention is high and

manage their panels. If Qualtrics respondents participate in more Web surveys than GfK panel members, these respondents may display higher levels of state dependence throughout the survey.

**Table 7** Cronbach's Alpha and Average Interitem Correlation for Material Success Items Across Experimental Conditions

	Position in the survey: Average interitem correlation (Cronbach's alphas)	
	Beginning	End
Items grouped together		
Keying sequence		
PK-NK-PK-NK	0.29 (0.62)	0.11 (0.32)
PK-PK-NK-NK	0.30 (0.63)	0.19 (0.48)
Items dispersed	0.47 (0.78)	0.39 (0.72)

Note. PK, positively keyed; NK, negatively keyed.

state dependence is low, respondents will notice a change in keying and adjust their response accordingly. However, at the end of the survey when attention is lower and state dependence higher, it is easier for respondents to display “misresponse” (Swain et al. 2008). Although the difference in reliabilities across keying sequences for grouped items at the end of the survey is substantial, it is not significant ( $\Pr(\alpha_{PK-NK-PK-NK} > \alpha_{PK-PK-NK-NK}) = 0.15$  for grouped items at the end of the survey) possibly because of the small number of observations per cell (37 respondents). We did not separate response by item keying when the scale items are dispersed because the filler items would affect the keying.

Thus, the model and the experimental results suggest reliability is especially poor when items of the balanced scale are grouped in an alternating fashion (PK-NK-PK-NK) at the end of the survey. Single-change keying (PK-PK-NK-NK) only slightly alleviates the problem. Even at the start of the survey, when attention is likely to be high, some state dependence is observed when items of an alternating keyed balanced scale are grouped together. The highest reliability occurs when the items of a balanced scale are dispersed at the beginning of the survey. Overall, our findings imply that if one uses a balanced scale with alternating keying sequence, it is preferable to disperse items and place them in the beginning of the survey (or have a shorter survey).

Naturally, there could be other factors that influence reliability and mitigate the effects found (Bradlow and Fitzsimons 2001, Schuman and Presser 1981, Sudman et al. 1996, Tourangeau et al. 2000). For example, to disperse items, we use two filler items between each focal scale's items, rather than dispersing items randomly throughout the survey. Consistent with Bradlow and Fitzsimons (2001), we believe that reliability would decrease if items were randomly dispersed across the entire survey. Similarly, using a scale consisting of only positively keyed items (PK-PK-PK-PK), we would expect that state dependence

would lead to the opposite result of (spuriously) high reliability when items are grouped together (Bradlow and Fitzsimons 2001). Thus, although high reliability is desirable, researchers should make sure that reliability is not artificially inflated by state dependence.

**State Dependence and Predictive Validity.** The reliability of a scale may not be indicative of external validity. For example, state dependence could produce an apparently highly reliable single keyed scale even when “true” reliability is poor. Accordingly, we examined whether item grouping and position in the survey affect predictive validity. If measurement is more accurate, the relationship between observed behavior and the scale score should be stronger. To test predictive ability, we investigate the correlation between the innovativeness scale scores from wave 1 of the study and the reported new product and service ownership collected in wave 2 of the study.

Table 8 reports the Cronbach's alphas and average interitem correlations from wave 1 as well as correlations between the innovativeness scale from wave 1 and product ownership reports from wave 2. First, let us examine the reliability results in the left side of each cell in Table 8. Consistent with the results of the material success scale in Table 7, *location* (beginning versus end) significantly affects reliability of the innovativeness scale ( $\Pr(\alpha_{end} > \alpha_{begin}) = 0.04$ ). Although directionally consistent with our theory, overall *grouping* does not have a significant main effect ( $\Pr(\alpha_{grouped} > \alpha_{dispersed}) = 0.17$ ). However, at the end of the survey, grouping does have a marginally significant effect ( $\Pr(\alpha_{grouped} > \alpha_{dispersed}) = 0.07$ ). As with the material success scale, measured reliability for the innovativeness scale is highest when state dependence is lowest at the beginning of the survey, when items are dispersed.

Next, let us examine predictive validity measures in the right side of each cell in Table 8. The innovativeness scale scores of respondents administered at the end of the survey in wave 1 are significantly less related to new product acquisition ( $r = 0.069$  for grouped items, and  $r = 0.090$  for dispersed items) than the scale scores of respondents who were administered the scale in the beginning of the survey ( $r =$

**Table 8** Cronbach's alpha and Average Interitem Correlation and Predictive Validity of the Innovativeness Scale Across Experimental Conditions

Item grouping	Position in the survey: Average interitem correlation (Cronbach's alphas)/correlations with new product ownership	
	Beginning	End
Items grouped	0.33 (0.75)/0.238	0.19 (0.59)/0.069
Items dispersed	0.43 (0.82)/0.243	0.29 (0.71)/0.090

0.238 for grouped items, and  $r = 0.243$  for dispersed items;  $\Pr(r_{end} > r_{begin}) = 0.04$ ). There is no main effect of grouping ( $\Pr(r_{grouped} > r_{dispersed}) = 0.29$ ), but the results suggest there may be a directional effect of grouping at the end of the survey.<sup>6</sup> The results suggest that predictive validity is highest when state dependence is lowest at the beginning of the survey and to a lesser extent when items are dispersed. It should be noted that the data used to assess validity were collected a week later and involve actual ownership of products compared with stated innovativeness, which may explain the relatively small correlations between the measured scale and actual behavior.

Overall, the findings for the innovativeness scale mostly replicate those for material success and suggest that for a balanced scale, both in terms of reliability and predictive validity, it is preferable to minimize state dependence by dispersing key items at the start of the survey.

## General Discussion

It is well known that content-irrelevant factors influence survey responses. Recent advances in computational power have enabled the use of increasingly sophisticated statistical models to identify and correct for systematic tendencies in survey responses. The present article builds on the response process literature by examining a systematic factor that may influence observed responses, state dependence—a form of response carryover from one item to another. The proposed psychometric model accounts for latent traits, individual-scale usage tendencies, and state dependence. The results demonstrate that state dependence significantly contributes to common method bias.

Although the state-dependence effect and scale usage heterogeneity substantially differ across countries, for all countries, respondents are most likely to be state dependent to extreme responses. Moreover, respondents are most likely to be state dependent to the answers they are inherently less likely to provide. We found an increase in state dependence at the end of a survey (with the exception of Japan), and longer

items generally lead to stronger state dependence. Changing item keying reduces state dependence.

Using an experimental design, we demonstrated that state dependence can affect Cronbach's alpha estimates and the predictive validity of constructs. The results suggest that state-dependence bias may be mitigated by placing measurement scales early in the survey, making items short, and avoiding grouping scale items together (by putting a few filler items in between the focal scale's items). Another important finding is that an exclusive focus on high reliability (by reviewers or journals) may not be good practice if reliability is spuriously high due to state dependence. For instance, the common practice of placing a grouped scale with only positively keyed items at the end of a long survey or experiment may produce an artificially high reliability. Of course, these findings are only indicative of the state-dependence effects and the need to pay attention to questionnaire design. The response process is complex, and many factors may influence it (Bradlow and Fitzsimons 2001, Schumann and Presser 1981, Sudman et al. 1996, Tourangeau et al. 2000).

## Limitations and Future Research

As in any article, the present research has limitations. Applying the statistical model requires a survey with at least one set of items with less than perfect absolute between-construct interitem correlations and a relatively large sample size (at least 400). Many marketing research studies, such as the one used in Study 1, satisfy these conditions. If the sample size is smaller, and interitem correlations are very high, the methodology is less useful. Some academic studies have a limited number of questions and respondents, making our proposed methodology less applicable to these studies. In such studies, it is especially important to focus on survey design rather than statistical techniques to ensure that state dependence will not seriously affect results.

Follow-up research could investigate which types of individuals and measures are most likely to be state dependent or suffer from state dependence as the survey progresses. Although this article explored variation in state dependence across nations and sociodemographics, it remains agnostic about the underlying psychological, motivational, and cultural factors that drive state dependence. In addition, this article does not explore questionnaires where overall evaluations might strongly influence the ratings of specific attributes (Beckwith and Lehmann 1975), presentation of items by subscales (Bradlow and Fitzsimons 2001), or common stimuli (Bradlow et al. 1999). Other statistical models might be needed to analyze such issues (e.g., Beckwith and Lehmann 1975).

An important issue is the generalizability of our findings to other scale formats. The statistical model is suited for both ordinal and nominal scales, which

<sup>6</sup> To see if we could replicate an effect of grouping on validity at the end of the survey, we collected additional data. We recruited 108 respondents randomly sampled from the U.S. adult population using the Qualtrics panel. Respondents were randomly assigned to two conditions: *grouped* innovativeness items at the end of the survey and *dispersed* innovativeness items at the end of the survey. At the start of each survey, well separated from the innovativeness scale, we asked the set of product ownership questions used in Study 2. The correlation between product ownership and innovativeness for grouped items at the end was 0.208, whereas the correlation between product ownership and innovativeness for dispersed items at the end was 0.353, which suggests a statistically significant grouping effect at the end of the survey.

are used most often in marketing (e.g., five-point or seven-point Likert scales, or choices on nominal scales). We also estimated a model in which we treat the Likert scale as an interval scale. For this purpose, a model is calibrated for the data in Study 2, replacing Equation (5) with

$$U_{i,k,c} = \beta_{i,c} + a_{k,c} \theta_{i,d(k)} + \gamma_{i,sd,k,c} (5 - |X_{i,k-1} - c|) + \varepsilon_{i,k,c}.$$

The interval scale model provided directionally similar results to our model, suggesting that our findings are robust, although its fit was slightly worse (LMD = -91,686 versus an LMD = -91,650 for our model). Ratio scales have a very large number of possible responses, making state dependence less likely. Future research could explore a state-dependence model for such question formats.

Little is known about whether the state-dependence effect can be reduced by training, making respondents aware of their tendencies, by incentives, or by changing response formats. Incentive alignment has been shown to improve external validity (Ding et al. 2005). Furthermore, future research could explore how different modes of administering the survey, sampling frames, and incentives influence state dependence.

A growing problem in online questionnaires is that respondents drop out of the survey, particularly for long questionnaires. Because the model estimates the degree of state dependence at the individual- and response-item levels, future research could investigate whether the proposed model can be used as a fatigue-detection tool to exclude respondents that reach overly high degrees of state dependence.

In summary, state dependence is an important yet understudied component of survey response. We hope that this article stimulates marketing researchers to pay more attention to state dependence in both domestic and international survey research.

### Electronic Companion

An electronic companion to this paper is available as part of the online version at <http://dx.doi.org/10.1287/mksc.1120.0722>.

### Acknowledgments

The authors thank Nicholas Reinoltz for providing research assistance for this project and seminar participants at Northwestern University, Singapore National University, the Free University Amsterdam, the (Marketing Science) conference in Cologne, and the Marketing Dynamics conference in Jaipur for helpful comments. The authors thank AiMark for providing the data in Study 1 and gratefully acknowledge financial support by the Erasmus Research Institute of Management and the Center for International Business Education and Research at Columbia Business School. Finally, M. G. de Jong thanks the Niels Stensen Foundation and The Netherlands Organization for Scientific Research for financial support.

### References

- Anderson JA (1984) Regression and ordered categorical variables. *J. Roy. Statist. Soc. Ser. B* 46(1):1–30.
- Ansari A, Jedidi K (2000) Bayesian factor analysis for multilevel binary observations. *Psychometrika* 65(4):475–496.
- Baumgartner H, Steenkamp J-BEM (1996) Exploratory consumer buying behavior: Conceptualization and measurement. *Internat. J. Res. Marketing* 13(2):121–137.
- Baumgartner H, Steenkamp J-BEM (2001) Response styles in marketing research: A cross-national investigation. *J. Marketing Res.* 38(2):143–156.
- Baumgartner H, Steenkamp J-BEM (2006) An extended paradigm for measurement analysis of marketing constructs applicable to panel data. *J. Marketing Res.* 43(3):431–442.
- Bearden WO, Netemeyer RG (1999) *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research*, 2nd ed. (Sage Publications, Newbury Park, CA).
- Bearden WO, Netemeyer RG, Teel JE (1989) Measurement of consumer susceptibility to interpersonal influence. *J. Consumer Res.* 15(4):473–481.
- Beckwith NE, Lehmann DR (1975) The importance of halo effects in multi-attribute attitude models. *J. Marketing Res.* 12(3):265–275.
- Benet-Martínez V, John OP (1998) Los Cinco Grandes across cultures and ethnic groups: Multitrait multimethod analyses of the big five in Spanish and English. *J. Personality Soc. Psych.* 75(3):729–750.
- Bickart B (1993) Carryover and backfire effects in marketing research. *J. Marketing Res.* 30(1):52–62.
- Bradlow ET, Fitzsimons GJ (2001) Subscale distance and item clustering effects in self-administered surveys: A new metric. *J. Marketing Res.* 38(2):254–261.
- Bradlow ET, Zaslavsky AM (1999) A hierarchical latent variable model for ordinal data from a customer satisfaction survey with “no answer” responses. *J. Amer. Statist. Assoc.* 94(445):43–52.
- Bradlow ET, Wainer H, Wang X (1999) A Bayesian random effects model for testlets. *Psychometrika* 64(2):153–168.
- Burisch M (1984) Approaches to personality inventory construction. *Amer. Psych.* 39(3):214–227.
- Chintagunta P (1998) Inertia and variety seeking in a model of brand-purchase timing. *Marketing Sci.* 17(3):253–270.
- Churchill GA Jr (1979) A paradigm for developing better measures of marketing constructs. *J. Marketing Res.* 16(1):64–73.
- Cronbach LJ (1946) Response set and test validity. *Ed. Psych. Measurement* 6:475–494.
- de Jong MG, Pieters R, Fox J-P (2010) Reducing social desirability via item-randomized response: An application to measure underreported desires. *J. Marketing Res.* 47(1):14–27.
- de Jong MG, Steenkamp J-BEM, Fox J-P, Baumgartner H (2008) Using item response theory to measure extreme response style in marketing research: A global investigation. *J. Marketing Res.* 45(1):104–115.
- DeSarbo WS, Lehmann DR, Hollman FG (2004) Modeling dynamic effects in repeated-measures experiments involving preference/choice: An illustration involving stated preference analysis. *Appl. Psych. Measurement* 28(3):186–209.
- Diener E, Emmons RA, Larsen RJ, Griffin S (1985) The satisfaction with life scale. *J. Personality Assessment* 49(1):71–75.
- Ding M, Grewal R, Liechty CJ (2005) Incentive-aligned conjoint analysis. *J. Marketing Res.* 42(1):67–82.
- Drolet A (2002) Inherent rule variability in consumer choice: Changing rules for change’s sake. *J. Consumer Res.* 29(3):293–305.
- Gal D, Rucker DD (2011) Answering the unasked question: Response substitution in consumer surveys. *J. Marketing Res.* 48(1):185–195.



- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statist. Sci.* 7(4):457–472.
- Gilks WR, Wild P (1992) Adaptive rejection sampling for Gibbs' sampling. *J. Roy. Statist. Soc. Ser. C* 41(2):337–348.
- Greenleaf EA (1992) Improving rating scale measures by detecting and correcting bias components in some response styles. *J. Marketing Res.* 29(2):176–188.
- Grice HP (1975) Logic and conversation. Cole P, Morgan JL, eds. *Syntax and Semantics*, Vol. 3 (Academic Press, New York), 41–58.
- Gruber RE, Lehmann DR (1983) The effect of omitting response tendency variables from regression models. Darden WR, Monroe KB, Dillon WR, eds. *Research Methods and Causal Modelling in Marketing* (American Marketing Association, Chicago), 131–136.
- Grunert SC, Juhl HJ (1995) Values, environmental attitudes, and buying of organic foods. *J. Econom. Psych.* 16(1):39–62.
- Guadagni PM, Little JDC (1983) A logit model of brand choice calibrated on scanner data. *Marketing Sci.* 2(3):203–238.
- Guilford JP (1954) *Psychometric Methods* (McGraw-Hill, London).
- Heckman J (1981) Heterogeneity and state dependence. Rosen S, ed. *Studies in Labor Markets National Bureau of Economic Research* (University of Chicago Press, Chicago), 91–139.
- Herzog RA, Bachman JG (1981) Effects of questionnaire length on response quality. *Public Opinion Quart.* 45(4):549–559.
- Holbrook MB, Schindler RM (1994) Age, sex, and attitude toward the past as predictors of consumers' aesthetic tastes for cultural products. *J. Marketing Res.* 31(3):412–422.
- Johnson MD, Lehmann DR, Horne DR (1990) The effects of fatigue on judgments of interproduct similarity. *Internat. J. Res. Marketing* 7(1):35–43.
- Jones MJ, Landwehr JT (1988) Removing heterogeneity bias from logit model estimation. *Marketing Sci.* 7(1):41–59.
- Kahn BE, Morrison DG, Wright GP (1986) Aggregating individual purchases to the household level. *Marketing Sci.* 5(3):260–268.
- Kahneman D, Knetsch JL, Thaler RH (1991) Anomalies—The endowment effect, loss aversion, and status-quo bias. *J. Econom. Perspect.* 5(1):193–206.
- Kamman R, Flett R (1983) Affectometer 2: A scale to measure current level of general happiness. *Australian J. Psych.* 35(2):259–265.
- Kane RB (1969) Computer generation of semantic differential (SD) questionnaires. *Ed. Psych. Measurement* 29(1):191–192.
- Keane MP (1997) Modeling heterogeneity and state dependence in consumer choice behavior. *J. Bus. Econom. Statist.* 15(3):310–327.
- Krosnick J (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. *Appl. Cognitive Psych.* 5(3):213–236.
- Levav J, Heitmann M, Herrmann A, Iyengar SS (2010) Order of product customization decisions: Evidence from field experiments. *J. Political Econom.* 118(2):274–299.
- Lichtenstein DR, Netemeyer RG, Burton S (1995) Assessing the domain specificity of deal proneness: A field study. *J. Consumer Res.* 22(3):314–326.
- Liechty JC, Fong DKH, DeSarbo WS (2005) Dynamic models incorporating individual heterogeneity: Utility evolution in conjoint analysis. *Marketing Sci.* 24(3):285–293.
- Lindell MK, Whitney DJ (2001) Accounting for common method variance in cross-sectional research designs. *J. Appl. Psych.* 81(1):114–21.
- McAlister L, Pessemier E (1982) Variety seeking behavior: An interdisciplinary review. *J. Consumer Res.* 9(3):311–322.
- Nunnally JC, Bernstein IH (1994) *Psychometric Theory* (McGraw-Hill, New York).
- Paap R, Franses PH (2000) A dynamic multinomial probit model for brand choice with different long-run and short-run effects of marketing-mix variables. *J. Appl. Econometrics* 15(6):717–744.
- Podsakoff PM, MacKenzie SB, Lee J-Y, Podsakoff NP (2003) Common method biases in behavioral research: A critical review of the literature and recommended remedies. *J. Appl. Psych.* 88(5):879–903.
- Richins ML, Dawson S (1992) A consumer values orientation for materialism and its measurement: Scale development and validation. *J. Consumer Res.* 19(3):303–316.
- Rindfleisch A, Malter AJ, Ganesan S, Moorman C (2008) Cross-sectional versus longitudinal survey research: Concepts, findings, and guidelines. *J. Marketing Res.* 45(3):261–279.
- Rossi PE, Gilula Z, Allenby GM (2001) Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *J. Amer. Statist. Assoc.* 96(453):20–31.
- Samuelson W, Zeckhauser R (1988) Status quo bias in decision making. *J. Risk Uncertainty* 1(1):7–59.
- Schuman H, Presser S (1981) *Questions and Answers in Attitude Surveys* (Academic Press, New York).
- Seetharaman PB, Ainslie A, Chintagunta PK (1999) Investigating household state dependence effects across categories. *J. Marketing Res.* 36(4):488–500.
- Shimp TA, Sharma S (1987) Consumer ethnocentrism: Construction and validation of the CETSCALE. *J. Marketing Res.* 24(3):280–289.
- Steenkamp J-BEM (1989) *Product Quality* (Van Gorcum, Assen, The Netherlands).
- Steenkamp J-BEM, Gielens K (2003) Consumer and market drivers of the trial rate of new consumer products. *J. Consumer Res.* 30(3):368–384.
- Steenkamp J-BEM, ter Hofstede F, Wedel M (1999) A cross-national investigation into the individual and cultural antecedents of consumer innovativeness. *J. Marketing* 63(2):55–69.
- Sudman S, Bradburn NM, Schwarz N (1996) *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology* (Jossey-Bass, San Francisco).
- Swain SD, Weathers D, Niedrich RW (2008) Assessing three sources of misresponse to reversed Likert items. *J. Marketing Res.* 45(1):116–131.
- Tourangeau R, Rips LJ, Rasinski K (2000) *The Psychology of Survey Response* (Cambridge University Press, Cambridge, UK).
- Van Rosmalen J, van Herk H, Groenen PJF (2010) Identifying response styles: A latent class bilinear multinomial logit model. *J. Marketing Res.* 47(1):157–172.
- Verhelst ND, Glas CAW (1993) A dynamic generalization of the rasch model. *Psychometrika* 58(3):395–415.
- Wainer H, Bradlow ET, Wang X (2007) *Testlet Response Theory and Its Applications* (Cambridge University Press, Cambridge, UK).
- Wang X, Bradlow ET, Wainer H (2002) A general Bayesian model for testlets: Theory and applications. *Appl. Psych. Measurement* 26(1):109–128.
- Wang X, Baldwin S, Wainer H, Bradlow ET, Reeve BB, Smith AW, Bellizzi KM, Baumgartner KB (2010) Using testlet response theory to analyze data from a survey of attitude change among breast cancer survivors. *Statist. Medicine* 29(19):2028–2044.
- Weijters B, Geuens M, Schillewaert N (2009) The proximity effect: The role of inter-item distance on reverse-item bias. *Internat. J. Res. Marketing* 26(1):2–12.
- Wong N, Rindfleisch A, Burroughs JE (2003) Do reverse-worded items confound measures in cross-cultural consumer research? The case of the material values scale. *J. Consumer Res.* 30(1):72–91.
- Yamamoto K (1995) Estimating the effects of test length and test time on parameter estimation using the HYBRID model. Technical Report TR-10, Educational Testing Service, Princeton, NJ.
- Zedeck S, Blood MR (1974) *Foundations of Behavioral Science Research in Organizations* (Brooks/Cole, Monterey, CA).