Customizing Persuasive Messages; the Value of Operative Measures

Author pre-print; please do not distribute.

Prof. Dr. Maurits C. Kaptein

Professor of Data Science & Health

Department of Statistics and Research Methods

Tilburg School of Social and Behavioral Sciences

Tilburg, the Netherlands.

&

Jheronimus Academy of Data Science (JADS), den Bosch, the Netherlands.

m.c.kaptein@uvt.nl

Abstract

Purpose – This article examines whether estimates of psychological traits obtained using meta-judgmental measures (as commonly present in CRM database systems) or operative measures are most useful in predicting customer behavior.

Design / methodology / approach – Using an online experiment ($N = 283$) we collect meta-judgmental and operative measures of customers. Subsequently, we compare the out-of-sample prediction error of responses to persuasive messages.

Finding – The paper shows that operative measures—derived directly from measures of customer behavior—are more informative than meta-judgmental measures.

Practical implications – Using interactive media it is possible to actively elicit operative measures. This article shows that practitioners seeking to customize their marketing communication should focus on obtaining such psychographic observations.

Originality / value – While currently both meta-judgmental measures and operative measures are used for customization in interactive marketing, this article directly compares their utility for predicting future consumer behavior.

Keywords. Consumer behavior, Integrated Marketing Communication, Persuasive Strategies, Measurement, Customization.

Customizing Persuasive Messages; the Value of Operative Measures

Author pre-print; please do not distribute.

Understanding the concepts and strategies behind integrated marketing communication (IMC) is important for both marketing academics and professionals (Peltier et al., 2003). IMC conceptualizes the *customization* of marketing communication: for example, Peltier et al. (2003) present a model conceptualizing how database management and IMC jointly operate to create customized communication campaigns. Their model provides a thorough overview of the relationships between consumer data, consumer profiles, and the resulting customized communication plan. However, the landscape in which IMC theory operates is continuously changing. To start with, the opportunities to dynamically interact with consumers are ever increasing: marketing communication is moving more and more towards a genuine dialogue (Bezjian-Avery et al., 1998; Haeckel, 1998; Deighton and Kornfeld, 2009; Finne and Grönroos, 2009). Also, the technological abilities to collect data are evolving: we have moved from a general inability to automatically measure consumer responses, to the use of electronic scanners, to a situation in which, in e-commerce for example, we can measure consumer responses to marketing communication continuously (see e.g., Parvinen et al., 2014; Kaptein et al., 2013; Kaptein and Parvinen, 2015). This opportunity seems like a treasure throve for IMC scholars: we can now truly reap the fruits of interactive IMC (Jiang and Chia, 2010). We can start to customize marketing communication not just at the level of (relatively large) customer segments, but we can, theoretically, move towards true one-to-one marketing (Stone and Woodcock, 2014) by measuring individual responses to marketing communication.

To effectively customize communication it is essential to have access to the right sources of data. Consumer data, often stored in customer databases (e.g., CRM data), is needed to profile consumers and to estimate the effects messages will have on customers. In essence, the key to customization is the ability to predict, *for each individual consumer*, which marketing message is most effective (Ansari and Mela, 2003; Hauser et al., 2009).

Despite an ever increasing access to data this prediction is still challenging, and although demonstrations of effective customization do exist (see, e.g., Ansari and Mela, 2003; Arora et al., 2008; Hauser et al., 2009), truly living up to the promises of one-to-one marketing seems outside the reach of many firms. Arguably, as advocated by Webster Jr. (1998), this inability might be due to the absence of psychographic measures such as motivation, needs, and attitudes in many (CRM-) databases; the marketeer must understand the psychological factors that motivate the observed behaviors for effective customization.

In this paper we focus on the opportunity that arises within interactive (online-) marketing communication to obtain psychological trait scores of consumers not by using survey instruments, but rather by logging the behavioral responses of consumers to carefully crafted marketing messages. Our contributions are to a) demonstrate in detail *how* we can quantify consumer behavior to obtain psychological trait scores by carefully designing our marketing messages, b) *illustrate this approach for a specific trait*, namely susceptibility to persuasive strategies (Kaptein and Eckles, 2012), that is highly valuable for customizing marketing communication (Cialdini, 2001; Kaptein and Eckles, 2012), and c) demonstrate that these trait scores not just *outperform classical survey measures* of the same trait scores—in terms of their ability to predict future consumer actions—but do so by an order of magnitude.

Theoretical background

In this section we introduce our background theory and develop our hypothesis. We first embed our work in the literature on database management, IMC, and customization. Next, we explain the difficulties that arise when measuring psychological traits of consumers and focus specifically on the distinction made in the psychological literature between *operative* and *meta-judgmental* measures of consumer traits; a nomenclature we use throughout. Finally, we discuss persuasive strategies as a prime subject for message customization and develop our hypotheses.

Database management and IMC

A distinguishing feature of marketing customization is the interplay between database managent and the interactive IMC plan (Peltier et al., 2003). Database management concerns data collection through surveys, internal records, etc., the curation of a database, and the process of segmenting or profiling customers. Next, an IMC plan can be created for each customer segment to customize campaigns (Peltier et al., 2003). Figure 1 highlights that multiple types of data can be of use in the customization process: a firm might have access to behavioral responses, demographics, and/or psychographic data. Each of these data types has its use, and likely the most effective customization can be attained by combining information from different sources (Kaptein and Parvinen, 2015). However, while demographics and behavioral responses that could reveal consumer preferences are often relatively easy to obtain, it is in practice often hard to obtain good psychographic measures (Webster Jr., 1998; Hauser et al., 2009; Kaptein and Parvinen, 2015).

Psychographic measures are historically collected using surveys, which are often expensive, impractical, or just outright impossible, to elicit from every customer. Because of this, and because of the relative ease with which behavioral responses can be collected by firms that operate largely online (e.g., interactive, web-based, communication), database curation seems to have shifted its focus from collecting psychographic measures to curating behavioral and tracking data (Finne and Grönroos, 2009). However, interactive technologies provide us not merely with the opportunity to track customers but also with an alternative method to obtain measures of psychological traits: by deliberately selecting a marketing message and observing the behavioral response, we can make inferences about psychological traits of consumers without the use of surveys (this opportunity is highlighted by the dashed recurrence arrow in Figure 1). To briefly illustrate, consider a consumers *cognitive style*: survey instruments to measure cognitive style have been actively developed (Hayes, 1998), but recently Hauser et al. (2009) used the observed response of customers to either textual or visually oriented messages to determine their cognitive style.

Note that in this work, by presenting customers with strategically chosen messages, the researchers were able to actively learn about their psychological make-up.

======================

Place Figure 1 about here

======================

Measuring consumer traits

Measuring psychological traits is notoriously hard, and a large literature on measuring traits exists both in marketing and in psychology (e.g., Geuens et al., 2009; Peter, 1979). There is a large body of work, under the heading of individual differences psychology (Dunton and Fazio, 1997; Furnham and Procter, 1989), that actively develops and evaluates measurement instruments for psychological traits. Examples include, but are not limited to, the measurement of personality (Gosling et al., 2003), emotional stability (Hills and Argyle, 2001), cognitive style (Huber, 1983), and locus of control (Ajzen, 2002). Many of these traits have found their use in marketing research, and also within marketing research itself trait scores and measurement instruments are actively developed (see, e.g., Geuens et al., 2009).

The measurement of psychological traits is complicated by the latent nature of traits; as opposed to (e.g.,) demographics, psychological traits can often not directly be observed and need to be deduced from a number of related measurements. A large literature on this topic exist, and in recent years scholars have begun to distinguish between formative and reflective measurement models (Cenfetelli and Bassellier, 2009): while the scores on reflective survey items are said to be caused by the latent construct, the formative model operates in the opposite direction: the measured items directly cause the latent score (Coltman et al., 2008). Another issue discussed in the literature is the method by which actual trait scores are constructed, these range from simple sum scores, to more elaborate linear combinations (factor scores), to the more recent concept of fuzzy sets, where a consumer is identified into a consumer segments using a score ranging between $0 - 1$ (Fiss, 2011). We refer the reader for a more thorough overview on the later topic to (Ragin, 2000).

The difficulties in measuring psychological traits demonstrate themselves not only by the large and diverse literature on the topic, but also manifest themselves through common concerns regarding validity and reliability. While many scales are internally consistent, the test-retest reliability of (e.g.,) personality traits is often low (Peter, 1979), and it is hard to develop novel instruments that are sufficiently reliable (Santor et al., 1997). Also the validity of many traits is debatable: often, the correlations with ostensibly related behaviors are small (Zanna et al., 1980). Also, the construct and external validity of many instruments is heavily debated (see, e.g., McCrae and Costa Jr, 1994). Hence, finding trait measures that are reliable and valid is non-trivial.

On the positive side, psychological traits, as opposed to behavioral observations or other more direct behavioral measures, are often believed to be context independent (Santor et al., 1997; McCrae and Costa Jr, 1994). Thus, many traits are considered stable—at least to some degree—over time and circumstances. This makes psychological trait scores potentially extremely valuable additions to CRM data: while consumer preferences might change rapidly over time, and while consumer interests in a specific product could fade abruptly after their search has ended, personality traits endure and can potentially be used to guide IMC customization in a multitude of situations and contexts.

A large body of work has focused specifically on de-contextualized measures of traits, favoring context independence over performance in specific situations (Furnham and Procter, 1989). Recently however, researcher both in marketing and psychology seem interested in method for measuring trait scores that might change over contexts at the benefit of performing better—in terms of predicting behavior—within the same context (Shobeiri et al., 2013).

Discussions of context dependence of traits have lead to a distinction in the psychological literature between so-called *operative* and *meta-judgmental* measures of a trait (Bassili, 1996). This distinction is best understood by borrowing from Polanyi (2012) and Jacoby and Kelley (1987) the distinction between psychological processes as “tools” and as “objects.” Psychological processes serve as tools when they are committed to accomplishing a task, such as in the decision to purchase a product. Psychological processes serve as objects when they are themselves the focus of attention, such as when respondents attempt to explain how they arrived at the decision to purchase a product. When filling out a survey, customers treat psychological processes as objects; they reflect on their own mental process. However, when responding to a marketing message, for example by clicking on a product, the psychological process is used as a tool to make the decision. Operative measures are likely to be more context dependent, since the behavior is manifest within a specific context, but at the added benefit of being more predictive of future behavior.

The current work focuses on differences in the ability to predict future consumer behavior based psychological traits obtained using either operative or meta-judgmental measures. Hence, it is worthwhile to detail how a firm could obtain scores using either measurement method. Suppose a firm intends to measure the extraversion (Gosling et al., 2003; Hirsh et al., 2012) of her customers. The meta-judgmental approach to obtaining these trait scores for all customers in a CRM system is as follows:

1. The firm finds a survey instrument to administer to its consumers.

2. The firm administers the survey. In practice, often only a small portion of customers in the CRM system is approached to fill out the survey, and trait scores for this sample are computed.

3. The firm generalizes the trait scores; if, for example, all the extraverts in the sample happen to be customers from a certain geographical region that bought a certain product offered by the firm we generalize and impute an “extravert” label to those customer who did not fill out the scale.

We contrast this method with the possible use of operative measures. Note that for a firm to be able to obtain operative measures it has to a) have recurrent interactions with its customers that allow the firm to experiment with different marketing messages, and b) be able to directly measure the behavioral response of a customer. If the firm has such opportunities obtaining operative measures is straightforward:

1. The firm designs (a minimum of) two versions of their messages; one appealing to introverts and one appealing to extraverts. Specifically, a product pitch appealing to either personal benefits or outside appearance could elicit different responses from introverts and extraverts.

2. The firm displays the messages randomly to her customers and monitors the behavioral responses; clicks on product pitches that stress personal benefits would lead to an increased introversion score, while clicks on products pitched using an appeal to outside appearance add to ones extraversion score.

3. As individual customers repeatedly interact with the firm, every interaction contributes to an updated score in the CRM system.

Note that to obtain operative measures we do not merely observe consumer behavior; rather, we specifically design our marketing messages to elicit behavioral responses that differ for consumers with different trait scores.

The importance of prediction for message customization

Before developing our hypotheses, we briefly examine the process of customization. Currently, the process of customization is often relatively static: we first collect customer information, and subsequently we profile customers (Schubert, 2000; Peltier et al., 2003) or estimate the effect of specific messages on segments (Kaptein and Parvinen, 2015). Next, we select a message for a distinct segment. Key to this process is the estimation step: we aim to select those messages which we deem most effective for the customer segment.

When using operative measures the process of customization becomes inherently *dynamic* since trait scores are derived from the observed behavior in previous interactions. The better the prediction of message success, the better the firm is able to select an effective future message (Kaptein and Parvinen, 2015).

Hypotheses

Based on the literature reviewed above, we now develop our hypotheses. First, we hypothesize the following

Hypothesis 1: *In an online selling context, stimuli can be created that elicit distint behavioral responses and allow a firm to compute (operative) trait-scores.*

This first hypothesis merely formalizes that the process as described above is feasible. We will examine this hypothesis by a) carrying out the process ourselves for a specific trait called “susceptibility to persuasive messages” (described below), and b) by demonstrating that we can replicate the finding in the literature that distinct customers respond differently to distinct persuasive messages (Kaptein and Eckles, 2012).

Second, in the psychological literature on meta-judgmental and operative measures it is well-known that participants find it cumbersome to reflect on their own psychological processes (Bassili, 1996); although meta-judgmental measures obtained using surveys are widespread, often the predictive validity of such measures is limited. One reason for the low predictive validity of meta-judgmental measures is their generality: as stated most meta-judgmental measures of consumer traits aim to quantify the trait irrespective of the context in which the trait manifests. Operative measures are obtained within a specific context; a consumer’s response to an advertisement appealing to extraverts can be used to quantify her extraversion score and is inherently tied to the context of responding to advertisements. This leads us to hypothesize the following:

Hypothesis 2: *Operative measures outperform meta-judgmental trait scores when used for predicting customer behavior in (a similar) context.*

Finally, we would like to stress the dynamic nature of operative measures; operative measures are refined and improved as more and more interactions with consumers accumulate. Hence, we formulate the following hypothesis:

Hypothesis 3: *The performance of operative measures increases as more and more responses are collected.*

In the current work we limit ourselves to a single trait; that of susceptibility to persuasive strategies. In the next section we detail the use of persuasive messages in IMC and describe why this trait can be of use for customizing IMC campaigns.

Customizing Persuasive Strategies

Researchers in consumer marketing and psychology have been studying the effects of persuasive messages (e.g. Cialdini and Trost, 1998; Cialdini and Goldstein, 2004), and numerous demonstrations of the effects of persuasive messages on people's attitudes (Crano and Prislin, 2006) and behaviors (Cialdini, 2005) exist. In this paper we examine the effect of customizing persuasive strategies using a subset of the strategies identified by (Cialdini, 2001). We focus on messages implementing the following strategies that are often encountered in online marketing:

- The *authority* strategy: When an authority figure tells people to do something, they typically do it (Milgram, 1974; Blass, 1991). Consumers are therefore frequently faced with authority endorsements of products such as "expert review".

- The *consensus* strategy: When individuals observe multiple others manifesting the same belief or behavior, they are more likely to believe and behave similarly (Ajzen and Fishbein, 1980; Goldstein et al., 2008; Zhu and Zhang, 2010).

- The *scarcity* strategy: Assumed scarcity increases the perceived value of products and opportunities (Cialdini, 2001), consequently advertisers and salespeople tend to use phrases such as "limited release", and "while supplies last"(Lynn, 1991).

Large average treatment effects of persuasive strategies have been found. However, researchers have also found evidence for heterogeneity in the effect of persuasive messages (Eagly, 1981; Haugtvedt and Petty, 1992; Cacioppo and Petty, 1982; Kaptein and Eckles, 2012; Chocarro et al., 2015). Hence, there is a clear indication that an interaction between psychological traits of consumers and their responses to persuasive strategies exists. We include the most prominent measures of these traits in our study:

- *Need for Cognition (NfC)*: a trait quantifying a person's willingness to think which is known to interact with the effectiveness of persuasive messages (Cacioppo and Petty, 1982).

- *Personality*: Hirsh et al. (2012) showed a direct relation of Big Five personality traits to responses to distinct persuasive appeals.

- *Susceptibility to persuasion*: Specifically for the strategies of authority, consensus and scarcity, Kaptein et al. (Kaptein et al., 2012) developed a survey instrument to measure an individuals tendency to comply to messages implementing these strategies.

Method

To test our hypotheses we measured consumer's responses to persuasive messages to obtain our operative measures, and we administered a number of questionnaires to obtain meta-judgmental measures. In the study participants were asked to fill out a survey and to state their Willingness to Pay (WtP) for nine different products. Each of the products was pitched using a different persuasive strategy (either "none", "authority", "social proof", or "scarcity").

Participants

Participants were $N = 283$ students enrolled in social science courses who participated for partial course credits (117 men, $M_{age} = 21.7$, $SD_{age} = 1.9$). The study was conducted fully online.

Procedure

Participants first filled out a survey that addressed demographics and subsequently measured several traits using standardized scales. Next, participants were sequentially presented with 9 advertisements each containing a cover image of a book, the name of the author, and the recommended price (see Figure 2). The books all fell in the same price range ($14 -$16) and subjects were all science-related.

Each book was accompanied with a "reason" why it was ostensibly selected that was displayed just above the book's cover image. These reasons implemented the consensus, scarcity and authority persuasive strategies. The consensus implementation was "World-wide bestseller!", the scarcity implementation was "Almost out of stock!" and the authority implementation was "The NYT recommends!". Each implementation was displayed twice or thrice (randomly) and there were two (neutral) books displayed *without* a persuasive message. The order of books, as well as the order of the labels and the matching of label and book were all randomized. After viewing the book, participants indicated how much they were willing to pay for each book in euros.

=======================

Place Figure 2 about here

=======================

Measures

Meta-judgmental Measures. We included a 30-item measure of the Big Five personality traits (Gosling et al., 2003); Extraversion (5 items, $\alpha = 0.76$), Conscientiousness (5 items, $\alpha = 0.91$), Agreeableness (5 items, $\alpha = 0.83$), Openness to Experience (5 items, $\alpha = 0.75$) and Neuroticism (5 items, $\alpha = 0.82$). For each of the five personality traits a composite score was computed by means of the factor scores (using

Principal Axis Factoring). We also administered the 18-item measure of Need for Cognition (NfC) ($\alpha = 0.82$). Next, participants were presented with the 30 items of the Susceptibility to Persuasion Scale (Kaptein et al., 2012). Factor scores were computed to obtain a scores for the susceptibility to Scarcity (5 items, $\alpha = 0.62$), Authority (4 items, $\alpha = 0.75$) and Consensus (4 items, $\alpha = 0.75$) sub-scales. Finally, participants were asked about their age, gender, ethnicity, living situation and academic major.

Operative Measures. The operative measures were derived from the willingness to pay (WtP) scores. We used a hierarchical model to estimate trait scores based on the measured WtP; details of this procedure and the model are presented in the analysis section. Here we provide some intuition: Table 1 gives the (hypothetical) responses of two distinct customers (A and B) to the books presented in the study. Displayed are both the persuasive strategy that accompanied the book, as well as the participants WtP. From these responses we can obtain a trait score: For example, for customer A, we first compute the participants average WtP for books that are presented without persuasive strategies (neutral): $(12 + 13)/2 = 12.5$. Next, we compute the average WtP for books presented with the authority strategy: $(15 + 16 + 18)/3 = 16\frac{1}{3}$. The difference between these two average WtP's provides an estimate for the participants trait score on "susceptibility to Authority": $16\frac{1}{3} - 12.5 = 3.83$.

======================

Place Table 1 about here

======================

The estimation procedure illustrated above is infeasible for several reasons: first, the obtained scores are not standardized, and second, the estimation of the averages based on only two or three observations is very noisy. These issues can be overcome by using an hierarchical model. In this model the noise in the individual level estimates is reduced by "borrowing strength" from other participants, and the scores are given in units of standard deviation(s) from the mean. Figure 3 shows a scatterplot of the (operative) trait scores obtained in our study.

=======================

Place Figure 3 about here

=======================

Analysis and results

The current study provided us with meta-judgmental measures of multiple traits and with a small set of operative measures. In this section we examine the utility of these measures for predicting future responses to marketing messages and formally test our hypotheses.

Manipulation check

Before comparing the two measurement methods it is of interest to see whether the persuasive strategies had any effect. Based on the literature we expect a positive average effect of using persuasive strategies. These effects are indeed found in our dataset: we find an average WtP of 10.93, ($SD = 7.57$) for the books presented without a persuasive strategy, while the average WtP for books presented using the scarcity, 11.13, ($SD = 7.92$) , authority, 11.56, ($SD = 8.74$) , and consensus, 12.45, ($SD = 8.93$) , strategies are all higher. The main effect of the use of persuasive messages is statistically significant, $F(3, 2563) = 3.87$, $p < .01$.

Computation of the predication error

We try to predict the response of participants when they evaluate their 9^{th} book using either a model of the relationships of the persuasive strategies interacting with the participants' personality traits, or using the participants previous responses to the books. For each participant $\texttt{id} = 1, \ldots, N = 283$ we have a number of meta-judgmental measures (e.g. `Openness`, and `Neuroticism`), and we have also have $\texttt{time}t = 1, \ldots, t = 9$ responses to products as observed over time. These responses are to $\texttt{Products} \in \{1, \ldots, 9\}$, presented with a $\texttt{Strategy} \in \{$"Neutral", "Social Proof", "Scarcity", "Authority"$\}$ and we observe the respondents willingness to pay (WtP).

Our aim is to predict the response of participants at $\texttt{Time} = 9$ *out of sample*: hence, we try to predict the WtP at $\texttt{Time} = 9$ for a specific participant without using this specific observation itself to fit the model. As the main comparison criterion we use the mean out of sample squared prediction error (OSE). OSE is defined as $\frac{1}{N} \sum (y_i^* - y_i)^2$, where y_i^* is the *predicted* WtP and y_i the actual observed WtP for respondent i on product 9.

For the evaluation of the OSE of the meta-judgmental measures we use a ten-fold cross validation approach (Hastie et al., 2013): we fit a model predicting $\texttt{WtP}|t = 8, \ldots$ using the first eight WtP ratings and all the traits present in the data of a randomly selected subset of 90% of our participants. We subsequently use the estimated coefficients from this model to predict $\texttt{WtP}|t = 9$ for the left-out 10% of participants and compute the OSE. We repeat this process $m = 100$ times.

To evaluate the OSE of the operative measures we take a slightly different approach: here we predict $\texttt{WtP}|t = 9$ using *only* participants previous responses $\texttt{WtP}|t = 1, \ldots, t = 8$. We fit a hierarchical model (see below) to estimate the effect of each strategy for each subject using the data of $t = 1, \ldots, t = 8$. This model *omits the meta-judgmental measures* and uses *only* the scores derived based on the previous responses. To enable direct comparison with the cross-validation procedure, the operative model(s) are also fit on a random selection of 90% of the participants of the original dataset. Subsequently, the OSE is computed over 10% of the dataset to provide comparable sampling variability.

Prediction Error using Meta-judgemental Measures. To predict the response to the 9^{th} product using meta judgmental measures we start by using a very elaborate model using all the traits measured in the study. In this model the outcome y_i of individual i is predicted using the following linear model:

$$
\begin{aligned}
y_i = &\beta_0 + \beta_{1,\dots} Strategy + \beta_{\dots} Gender \times Strategy + \beta_{\dots} Age \times Strategy + \\
&\beta_{\dots} Extraversion \times Strategy + \beta_{\dots} Conscientiousness \times Strategy + \\
&\beta_{\dots} Agreeableness \times Strategy + \\
&\beta_{\dots} Openness \times Strategy + \beta_{\dots} Neuroticism \times Strategy + \\
&\beta_{\dots} NfC \times Strategy + \beta_{\dots} Scarcity \times Strategy + \beta_{\dots} Authority \times Strategy + \\
&\beta_{\dots,k} Concensus \times Strategy + \\
&+ \lambda_{j[i]} + \nabla_{l[i]} + \epsilon
\end{aligned}
$$

where $\beta_{\dots}$ stands for the estimated coefficient of the effect (which are many, given the dummy coding of `Strategy`), $\lambda_j \sim \mathcal{N}(0, \sigma_j^2)$ and is a random effect of `Strategy`, $\nabla_l \sim \mathcal{N}(0, \sigma_l^2)$ and is the random effect of `Product`, and $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$, the error variance. With the interactions terms denoted using "$\times$" we mean that we include both the interaction as well as the associated main effects. The terms "Scarcity", "Authority", and "Consensus" refer to the scores obtained on the susceptibility to persuasion scale. We refer to this full model as *Model 1.* We use Model 1 as the base model to predict `WtP`$|t = 9$. However, due to its large number of parameters this model likely overfits the data (Cf. Gaber et al., 2005). Therefore, we include two alternatives: First, we examine a model in which we use a series of model comparisons and select the model with the lowest BIC: we select this model automatically using `dredge` from the `[R]` package `MuMIn`. The selected

Model 2 is:

$$
\begin{aligned}
y_i = &\beta_0 + \beta_{1,\ldots} Strategy + \\
&\beta_{\ldots}(Extraversion \times Strategy) + \\
&\beta_{\ldots} Agreeableness + \\
&\beta_{\ldots} NfC + \\
&\beta_{\ldots} Scarcity \times Strategy + \\
&\beta_{\ldots,k} Concensus + \\
&\lambda_{j[i]} + \nabla_{l[i]} + \epsilon
\end{aligned}
$$

Second, we examine the use of regularization (see Hastie et al., 2013, for a discussion). For this approach we use the random effects Lasso as defined in `GLMMLasso` to subset the number of predictors. Since the `GLMLasso` selects a different *Model 3* for each of the 100 iterations we do not write out the models explicitly. Note that in each of these models the average behavior of customers is included through the random effects terms; hence these models effectively contain demographic data, psychographic data, *and* historical behavioral data to predict the next response.

Prediction Error using Operative Measures. We use $\texttt{WtP}|t = 1, \ldots, t = 8$ and the structure of the messages to obtain scores on the operative measures and predict $\texttt{WtP}|t = 9$. The model that we use is:

$$y_i = \beta_{j[i]} X_{j[i]} + \nabla_{l[i]} + \epsilon \tag{1}$$

In this specification $\beta_{j[i]}$ is a vector containing the estimated effects for each of the persuasive strategies where $\beta_{j[i]} \sim \mathcal{MVN}(\beta_0, \Sigma)$. Thus, there is a random effect (over participants) for each strategy. The Maximum a Postriori (MAP) estimates of the individual level coefficients of the random effects are regarded the operative trait scores. As above, $\nabla_{l[i]}$ is the random effect for the products. Here y_i contains $\texttt{WtP}|t = 1, \ldots, t = 8$ (and excludes $\texttt{WtP}|t = 9$).

Testing of our hypotheses

Our first hypothesis concerns the ability to create operative measures. While Figure 3 already displayed the estimated operative measures for consumers included in our study, this Figure does not provide a formal test of the ability to quantify meaningful variance in consumer traits. Such a test can be provided by looking at the variance components of the model described in Eq. 1 (see also Kaptein and Eckles, 2012).

Table 2 displays the estimated fixed effects and random effects obtained when fitting the model described in Eq. 1 to our full dataset. From the Table it is clear—as in the manipulation check—that we find positive average effects of the use of persuasive messages. However, it is also clear that the estimated random effects are (comparatively) large; the estimated standard deviations larger than the main effects. In effect, for a large number of participants the estimated effect of certain strategies turns out to be negative. A formal test comparing a model without the random terms for the strategies with the larger model including random strategy variation rejects the null-hypothesis of zero variance components, $p < .001$, $\chi^2 = 113.09$ and confirms hypothesis one.

======================

Place Table 2 about here

======================

To test our second hypothesis, Table 3 present the performance in terms of OSE of each of the different models. It is clear that when using all the data the OSE of the model that uses operative measures is (magnitudes) lower then that of any of the three models using only meta-judgmental measures ($p < .05$ in each case using pairwise t-tests with Bonferroni corrections). This confirms our hypothesis.

======================

Place Table 3 about here

======================

To examine hypothesis three, we compare the performance of the best performing meta-judgmental model (Model 3) with the performance of the operative measures model as more and more data accumulates. Table 4 shows the obtained difference in OSE between the use of operative measures based on $t = 3, \ldots, t = 8$ observations and meta-judgmental measures. We see that the difference in model performance increases as more operative measures are available. Hence, as hypothesized, more and more observations lead to better and better trait scores.

=======================

Place Table 4 about here

=======================

To conclude, operative measures outperformed the meta-judgmental measures in our current context; this supports our hypotheses. We distribute the annotated dataset resulting from our experiment with this paper to allow others to improve on our prediction(s) using meta-judgemental measures. The dataset can be retrieved at `http://dx.doi.org/10.7910/DVN/OXDAE9`.

Discussion

In this paper we examined the use of operative measures versus meta-judgmental measures for use in IMC customization. We have shown that operative measures can outperform meta-judgmental measures when it comes to predicting future consumer behavior. Although, based on the current study, this result might be specific only to customizing persuasive messages, we content to have shown the potential of using operative measures in (online) marketing communication: a well-chosen selection of marketing messages can be used to obtain psychographic measures of consumers traits.

Theoretical implications

Our study has a number of theoretical implications. First of all, the stellar performance of operative measures raises a number of questions regarding validity and reliability of both meta-judgmental measures and operative measures. In our study, the meta-judgmental measures seemed to have a low predictive validity. However, a number of these measures, such as the Big Five, have previously shown to be reliable and valid in other settings (Santor et al., 1997; Gosling et al., 2003). Perhaps our field has focused too much on internal consistence, test-retest reliability, and construct validity as opposed to predictive validity? On the other hand, operative measures beg further scrutiny; high predictive validity is an asset but the test re-test reliability and the construct validity of operative measures remain open questions.

Another theoretical challenge is highlighted by our first hypothesis; while we managed to create relatively simple marketing messages that elicit distinct responses from consumers, theories motivating different psychological traits often do not link one-to-one with possible marketing messages. For a large number of often-used psychological traits their operationalization in marketing messages is not clear-cut. It is thus a theoretical challenge to move from the survey items now used to measure traits, to marketing messages that would allow operative measurement. Our study also raises a theoretical question regarding the interactions between multiple traits: when collecting meta-judgmental measures, often a large number of items are administered to consumers, while the number of observations involved in constructing an operative measure is inherently limited by the number of interactions. Because of this limitation, it becomes interesting to see whether multiple traits can be queried using a single marketing message.

The distinction between operative and meta-judgmental measures raises questions regarding the context dependency of trait scores. Traditionally, psychologist have searched for context independent trait scores in the sense that these traits are, as much as possible, a property of the person as opposed to a person-environment combination. However, no binary distinction between the two exists; to a lesser or greater extend the environment will have an impact on the manifestations of a trait. The current work thus highlights a theoretical challenge of creating operationalizations of context and context dependency that are useful for IMC: one could argue that we are looking for trait scores that inform us about consumer responses to marketing messages, not for trait scores that are possibly useful in a broader context but provide less clarity in our own domain. We need to theoretically understand this trade-off.

Finally, operative measures, since they are often obtained implicitly (Kaptein et al., 2015), also raise a number of questions regarding the perception of privacy of consumers and their trust in the firm. While customers have, in recent years, become aware of the fact that their product preferences are derived from their browsing behavior, it remains to be seen whether deriving psychological traits form actively elicited behavior is feasible.

Managerial implications

Next to the theoretical implications, our study potentially impacts IMC practice. At minimum, we have demonstrated an alternative method—as opposed to the standard survey practice—by which CRM systems can be enriched with psychological trait scores. The great predictive performance of these scores should highlight the possible importance of our method for (online) IMC customization. However, the costs and efforts involved in obtaining operative measures will differ tremendously from firm to firm. For firms who operate largely online, have a habit of (e.g.,) AB testing versions of their webpages or marketing messages, and already have the ability to store and process behavioral responses, obtaining operative traits measures boils down to a fairly limited effort of consciously designing marketing messages such that these elicit different responses from different customers. On the other hand, for those companies who do not have these facilities in place, setting up such an infrastructure might be much more costly than obtaining meta-judgmental measures for a sample of customers and imputing scores to the remaining customers in their CRM database.

The practice of obtaining operative measures raises practical issues regarding privacy and trust. The practice of imputing psychological trait scores based on behavioral traces (such as carried out by Kosinski et al., 2013) is heavily debated in the popular media. We encourage firms to be transparent regarding their practices in this regard, and would encourage an open dialogue with customers. In the end, IMC customization aims to support long-term customer satisfaction, not short-term monetary gains. Firms should not, in the process, ignore the possible costs in terms of privacy and trust that customers might experience when confronted with operative trait measurements.

Our results imply a possible change of focus for many interactive marketing practitioners: Many companies currently have CRM systems containing rich descriptions of their customers gender, age, background, social status, etc. etc. Possibly, these CRM databases also contain psychographic data, although these remain hard to obtain: it is often practically infeasible to administer large surveys to customers and the subsequent imputations remain noisy. However, we have shown that a relatively small number behavioral responses to well-chosen messages can be used to estimate psychological traits and subsequently estimate message effectiveness. These estimates can subsequently be used to customize content.

Limitations and future work

Despite clearly highlighting the potential of operative measures, our work has a number of limitations. First of all, we feel the use of willingness to pay—as opposed to actual payments—needs further scrutiny: our study should be replicated outside the laboratory setting using actual behavioral outcomes. Second, our study seemingly benefitted the performance of operative measures by using the operative trait scores to predict future customer behavior in the exact same context. We have already discussed the theoretical importance of further investigating the role of context in this setting, but would like to see replications of our work in which the performance of operative and meta-judgmental trait scores is investigated for increasingly more distinct contexts: our study could be replicated exploring consumer responses in a different online store, for a different product category, or, in the extreme case, even outside the marketing context.

Finally, future work needs to create the actual stimuli that elicit behavioral responses that allow firms to infer trait scores. How exactly would we create (e.g.,) emailed marketing messages such that they appeal differently for customers with (e.g.,) different level of agreeableness? Furthermore, if we manage to design such messages, how can we judge the construct validity and reliability of the resulting trait scores? We hope the current paper encourages future work in this direction.

References

Ajzen, I. (2002), “Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior1”, *Journal of Applied Social Psychology* , Vol. 32 No. 4, pp. 665–683.

Ajzen, I. and Fishbein, M. (1980), *Understanding attitudes and predicting social behavior*, Prentice-Hall, Englewood Cliffs, NJ.

Ansari, A. and Mela, C. F. (2003), “E-Customization”, *Journal of Marketing Research*, Vol. 40 No. 2, pp. 131–145.

Arora, N., Dreze, X., Ghose, A., Hess, J. D., Iyengar, R., Jing, B., Joshi, Y., Kumar, V., Lurie, N. and Neslin, S. (2008), “Putting one-to-one marketing to work: Personalization, customization, and choice”, *Marketing Letters*, Vol. 19 No. 3, pp. 305–321.

Bassili, J. N. (1996), “Meta-judgmental versus operative indexes of psychological attributes: The case of measures of attitude strength”, *Journal of Personality and Social Psychology*, Vol. 71 No. 1, pp. 637–653.

Bezjian-Avery, A., Calder, B. and Iacobucci, D. (1998), “New media interactive advertising vs. traditional advertising”, *Journal of Advertising Research*, Vol. 38 No. 4, pp. 23–32.

Blass, T. (1991), “Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions.”, *Journal of Personality and Social Psychology*, Vol. 60 No. 3, pp. 398–413.

Cacioppo, J. T. and Petty, R. E. (1982), “The need for Cognition.”, *Journal of Personality and Social Psychology.*, Vol. 42 No. 1, pp. 116–131.

Cenfetelli, R. T. and Bassellier, G. (2009), “Interpretation of formative measurement in information systems research”, *Mis Quarterly*, JSTOR, pp. 689–707.

Chocarro, R., Cortiñas, M. and Villanueva, M. L. (2015), "Customer heterogenity in the development of e-loyalty", *Journal of Research in Interactive Marketing*, Vol. 9 No. 3, pp. 190–213.

Cialdini, R. (2001), *Influence, Science and Practice*, Allyn & Bacon, Boston.

Cialdini, R. B. (2005), "Don't Throw in the Towel: Use Social Influence Research", *American Psychological Society Observer*, Vol. 18 No. 4, pp. 33–34.

Cialdini, R. B. and Goldstein, N. J. (2004), "Social influence: compliance and conformity.", *Annual Review of Psychology*, Vol. 55 No. 1, Annual Reviews, pp. 591–621.

Cialdini, R. B. and Trost, M. R. (1998), "Social Influence: Social Norms, Conformity, and Compliance", in D. T. Gilbert, S. T. Fiske and G. Lindzy (eds), *The Handbook of Social Psychology*, Vol. 2, McGraw-Hill, chapter 21, pp. 151–192.

Coltman, T., Devinney, T. M., Midgley, D. F. and Venaik, S. (2008), "Formative versus reflective measurement models: Two applications of formative measurement", *Journal of Business Research*, Vol. 61 No. 12, pp. 1250–1262.

Crano, W. D. and Prislin, R. (2006), "Attitudes and persuasion.", *Annual Review of Psychology*, Vol. 57 No. 1, pp. 345–74.

Deighton, J. and Kornfeld, L. (2009), "Interactivity's Unanticipated Consequences for Marketers and Marketing", *Journal of Interactive Marketing*, Vol. 23 No. 1, pp. 4–10.

Dunton, B. C. and Fazio, R. H. (1997), "An individual difference measure of motivation to control prejudiced reactions", *Personality and Social Psychology Bulletin*, Vol. 23 No. 3, pp. 316–326.

Eagly, A. H. (1981), "Recipient characteristics as determinants of responses to persuasion", in R. E. Petty, T. M. Ostrom and T. C. Brock (eds), *Cognitive Responses in Persuasion*, Lawrence Erlbaum, pp. 173–195.

Finne, Å. and Grönroos, C. (2009), "Rethinking marketing communication: From integrated marketing communication to relationship communication", *Journal of Marketing Communications*, Vol. 15 No. 1, pp. 179–195.

Fiss, P. C. (2011), "Building better causal theories: A fuzzy set approach to typologies in organization research", *Academy of Management Journal*, Vol. 54 No. 2, pp. 393–420.

Fogg, B. J. (2009), "The Behavior Grid: 35 ways behavior can change", in S. Chatterjee and P. Dev (eds), *Proceedings of the Fourth International Conference on Persuasive Technology*, ACM, New York, NN pp. 42.

Furnham, A. and Procter, E. (1989), "Belief in a just world: Review and critique of the individual difference literature", *British Journal of Social Psychology*, Vol. 28 No. 4, pp. 365–384.

Gaber, M. M., Zaslavsky, A. and Krishnaswamy, S. (2005), "Mining data streams", *ACM SIGMOD Record*, Vol. 34 No. 2, ACM, pp. 18.

Geuens, M., Weijters, B. and De Wulf, K. (2009), "A new measure of brand personality", *International Journal of Research in Marketing*, Vol. 26 No. 2, pp. 97–107.

Goldstein, N. J., Cialdini, R. B. and Griskevicius, V. (2008), "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels.", *Journal of Consumer Research*, Vol. 35 No. 3, pp. 472–482.

Gosling, S. D., Rentfrow, P. J. and Swann W. B., J. (2003), "A very brief measure of the Big-Five personality domains", *Journal of Research in Personality*, Vol. 37 No. 1, pp. 504–528.

Haeckel, S. H. (1998), "About the nature and future of interactive marketing", *Journal of Interactive Marketing*, Vol. 12 No. 2, pp. 63–71.

Hastie, T., Tibshirani, R. and Friedman, J. (2013), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Vol. 11, Springer Science & Business Media.

Haugtvedt, C. P. and Petty, R. E. (1992), "Personality and persuasion: Need for cognition moderates the persistence and resistance of attitude changes.", *Journal of Personality and Social Psychology*, Vol. 63 No. 2, pp. 308–319.

Hauser, J. R., Urban, G. L., Liberali, G. and Braun, M. (2009), "Website Morphing", *Marketing Science*, Vol. 28 No. 3, pp. 202–223.

Hayes, J. (1998), "Cognitive Style and the Theory and Practice of Individual and Collective Learning in Organizations", *Human Relations*, Vol. 51 No. 7, pp. 847–871.

Hills, P. and Argyle, M. (2001), "Emotional stability as a major dimension of happiness", *Personality and individual differences*, Vol. 31 No. 8, pp. 1357–1364.

Hirsh, J. B., Kang, S. K. and Bodenhausen, G. V. (2012), "Personalized Persuasion: Tailoring Persuasive Appeals to Recipients' Personality Traits", *Psychological Science*, Vol. 23 No. 4, pp. 578–581.

Huber, G. P. (1983), "Cognitive style as a basis for mis and dss designs: Much ado about nothing?", *Management Science*, Vol. 29 No. 5, pp. 567–579.

Jacoby, L. L. and Kelley, C. M. (1987), "Unconscious Influences of Memory for a Prior Event", *Personality and Social Psychology Bulletin*, Vol. 13 No. 3, pp. 314–336.

Jiang, P. and Chia, S. L. (2010), "Developing Integrated Marketing Communication IMC in online communities: A conceptual perspective from Search, Experience and Credence segmentation", *International Journal of Internet Marketing and Advertising*, Vol. 6 No. 1, pp. 22.

Kaptein, M. C., de Ruyter, B., Markopoulos, P. and Aarts, E. (2012), "Tailored Persuasive

Text Messages to Reduce Snacking.", *Transactions on Interactive Intelligent Systems*, Vol. 2 No. 2, pp. 10–35.

Kaptein, M. C. and Eckles, D. (2012), "Heterogeneity in the Effects of Online Persuasion", *Journal of Interactive Marketing*, Vol. 26 No. 3, pp. 176–188.

Kaptein, M., Markopoulos, P., De Ruyter, B. and Aarts, E. (2015), "Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles", *International Journal of Human-Computer Studies*, Vol. 77, pp. 38–51.

Kaptein, M. and Parvinen, P. (2015), "Advancing E-Commerce Personalization: Process Framework and Case Study", *International Journal of Electronic Commerce*, Vol. 19 No. 1, pp. 7–33.

Kaptein, M., Parvinen, P. and Poyry, E. (2013), "Theory vs. Data-Driven Learning in Future E-Commerce", in *Proceedings of the 2013 46th Hawaii International Conference on System Sciences (HICSS)*, pp. 2763–2772.

Kosinski, M., Stillwell, D. and Graepel, T. (2013), "Private traits and attributes are predictable from digital records of human behavior", *Proceedings of the National Academy of Sciences*, Vol. 110 No. 15, pp. 5802–5805.

Lynn, M. (1991), "Scarcity effects on value: A quantitative review of the commodity theory literature", *Psychology and Marketing*, Vol. 8 No. 1, pp. 43–57.

McCrae, R. R. and Costa Jr, P. T. (1994), "The stability of personality: Observations and evaluations", *Current Directions in Psychological Science*, Vol. 3 No. 6, pp. 173–175.

Milgram, S. (1974), *Obedience to Authority*, Tavistock, London.

Parvinen, P., Oinas-Kukkonen, H. and Kaptein, M. (2014), "E-selling: A new avenue of research for service design and online engagement", *Electronic Commerce Research and Applications*, Vol. 14 No. 4, pp. 214–221.

Peltier, J. W., Schibrowsky, J. A. and Schultz, D. E. (2003), “Interactive integrated marketing communication: Combining the power of IMC, the new media and database marketing”, *International Journal of Advertising*, Vol. 22 No. 1, pp. 93–115.

Peter, J. P. (1979), “Reliability: A review of psychometric basics and recent marketing practices”, *Journal of marketing research*, Vol. 16 No. 1, pp. 6–17.

Polanyi, M. (2012), *Personal Knowledge: Towards a Post-Critical Philosophy*, University of Chicago Press, Chicago.

Ragin, C. C. (2000), *Fuzzy-set social science*, University of Chicago Press, Chicago.

Santor, D. A., Bagby, R. M. and Joffe, R. T. (1997), “Evaluating stability and change in personality and depression.”, *Journal of Personality and Social Psychology*, Vol. 73 No. 6, pp. 1354.

Schubert, P. (2000), “The Participatory Electronic Product Catalog: Supporting Customer Collaboration in E-Commerce Applications”, *Electronic Markets*, Vol. 10 No. 1, pp. 229–236.

Shobeiri, S., Laroche, M. and Mazaheri, E. (2013), “Shaping e-retailer’s website personality: The importance of experiential marketing”, *Journal of Retailing and Consumer Services*, Vol. 20 No. 1, pp. 102–110.

Stone, M. D. and Woodcock, N. D. (2014), “Interactive, direct and digital marketing - a future that depends on better use of business intelligence”, *Journal of Research in Interactive Marketing*, Vol. 8 No. 2, pp. 12.

Webster Jr., F. E. (1998), “Interactivity and marketing paradigm shifts”, *Journal of Interactive Marketing*, Vol. 12 No. 1, pp. 54–55.

Zanna, M. P., Olson, J. M. and Fazio, R. H. (1980), “Attitude–behavior consistency: An individual difference perspective.", *Journal of Personality and Social Psychology*, Vol. 38 No. 3, pp. 432.

Zhu, F. and Zhang, X. M. (2010), "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics", *Journal of Marketing*, Vol. 74 No. 2, pp. 133–148.